National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

NL-339 (r. 88/04)

Canada

THE UNIVERSITY OF ALBERTA

PATTERN REPRESENTATION AND CLASSIFICATION OF THE EEG USING THE KARHUNEN-LOEVE TRANSFORM

BY

© MICHAEL STANLEY LAZAR

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND
RESEARCH IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF ELECTRICAL ENGINEERING

EDMONTON, ALBERTA
FALL, 1988

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-45608-6

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR : Michael Stanley Lazar

TITLE OF THESIS : Pattern Representation and Classification of the EEG Using The Karhunen-Loeve Transform

DEGREE : Master of Science

YEAR THIS DEGREE GRANTED : 1988

Permission is hereby granted to **THE UNIVERSITY OF ALBERTA LIBRARY** to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

Michael Lazar

16111 - 88A Avenue
Edmonton, Alberta
Canada
T5R 4N5

Date : July 27, 1988

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Pattern Representation and Classification of the EEG Using the Karhunen-Loeve Transform submitted by Michael Stanley Lazar in partial fulfilment of the requirements for the degree of Master of Science

Z. J. Koles
Supervisor

R. E. Rink

Xiaobo Li

J. N. McMullin

Date : July 20, 1988

Dedicated with much love to my
mother, brother, and late father.

ABSTRACT

In this thesis the Karhunen-Loeve transform (KLT), also known as Principal Components Analysis (PCA), has been used in pattern representation and classification of temporal sequences of spatial EEG maps. The KLT allows the underlying structure of a set of data to be examined by transforming the input data into a set of orthogonal basis functions and a set of orthogonal coefficients which weight the basis functions. The KLT was applied to two classes of EEG data, the eyes open resting (EP) and eyes closed resting (EC) states. Representational accuracy was measured in terms of the amount of power (variance) accounted for by the basis images and associated coefficients, and how well the reconstructed images correlated with the original images.

It was found that the first 5 basis images typically accounted for over 90% of the variance in EEG segments ranging in length from 0.5 to 40 seconds, and that the reconstructed images correlated highly on average with the original images. It was also found that basis images computed from one cognitive state better represented EEG from the same state than from another state, even for inter-subject data.

The second set of analyses used a transformation based on the KLT and developed by Fukunaga and Koontz (1970) to generate parameters and features amenable for classification. The features were then used in conjunction with a minimum distance classifier to differentiate between EC and EP data. Accurate supervised classification was possible for EEG data when training vectors from the subject to be classified were available, regardless of whose data was used in computing the classification parameters.

ACKNOWLEDGEMENTS

I would like to gratefully acknowledge the Natural Sciences and Engineering Research Council of Canada for providing me the financial support to pursue this degree. I would also like to thank the Alberta Heritage Foundation for Medical Research for a research grant, which has given me the opportunity to travel and defer the costs of compiling this thesis.

Within the Department of Applied Sciences in Medicine there are many who deserve considerable accolades. To my supervisor, Dr. Z. Koles, my sincerest thanks for giving me direction and valuable insight and yet enough free reign to learn a great deal on my own. To Bob Morse, my mentor of the last two years, I owe not only thanks for help in my research but also for becoming a friend and enriching me personally through his wit and charm, something which I shall always cherish. I also owe Raman Paranjape special mention for the many interesting and valuable conversations we have held regarding both our work and lives. To Dr. Brendon Maclean, Bob Heath, Heather Lenz, Bob Arthur, Dr. Snyder, Dr. Kasmia, Satyapal Rathee, Anthony Soong, Ion Buicliu, Elias Haska, Narc Ouellette, Dr. Overton, Dr. Fenna, and all of the other fine people in ASM, go my sincerest thanks and best wishes.

For my life outside the University I would like to acknowledge the support of my family and many friends. Above all my mother for providing me with love and a home, and my brother for the many stimulating conversations and games of raquetball. Although there are too many others to list extra special thanks go to Rey Ledda, Ronald Tychsen, Steve Knudsen and especially Rita J. Kolpak for providing me with great EEG data and a wonderful time in my life.

TABLE OF CONTENTS

Chapter 1 The Electroencephalogram..........1
1.0 Introduction..........1
1.1 Origin of Scalp Voltage Potentials..........4

Chapter 2 The Karhunen-Loeve Transform..........8
2.0 Introduction..........8
2.1 Data Transformation..........8
    2.1.1 Basic Notation of Linear Transformations..........9
2.2 Pattern Recognition..........14
2.3 Karhunen-Loeve Transform Theory..........17
    2.3.1 Review of Eigen Analysis..........18
    2.3.2 Autocorrelation Functions..........19
    2.3.3 Mathematical Development of the Karhunen-Loeve Transform..........20
    2.3.4 Additional Properties and Remarks..........25

Chapter 3 The Karhunen-Loeve Transform Applied to Pattern Representation of The Spatial EEG..........30
3.0 Introduction..........30
3.1 Data Acquisition and Experimental Set Up..........31
3.2 Analyzing Image Sequences with the KLT..........33
3.3 Program Notes..........35
3.4 Pattern Representation of the EEG Using the Karhunen-Loeve Transform Within Subject Records..........38
    3.4.1 Detailed Study - Methods & Results..........44
    3.4.2 Detailed Study - Discussion..........59
    3.4.3 Multiple Subject Study - Methods & Results..........66
    3.4.4 Multiple Subject Study - Discussion..........70
3.5 Pattern Representation of the EEG Using the Karhunen-Loeve Transform on Subject Groups..........73
    3.5.1 Pattern Representation Across Cognitive States Within and Individual - Methods and Results..........75
    3.5.2 Pattern Representation Across Cognitive States Within an Individual - Discussion..........77
    3.5.3 Pattern Representation Across Individuals in the Same State - Methods and Results..........80
    3.5.4 Pattern Representation Across Individuals in the

Same State - Discussion 82
3.5.5 Pattern Representation Using Collective Data - Methods and Results 83
3.5.6 Pattern Representation Using Collective Data - Discussion 87
3.6 Discussion and Conclusions 91

Chapter 4 Pattern Classification of the EEG Using the Karhunen-Loeve Transform 96
4.0 Introduction 96
4.1 The Fukunaga-Koontz Transformation 98
4.2 Classification on Individual Data - Methods and Results 102
4.3 Classification on Individual Data - Discussion 106
4.4 Classification Across Individuals - Methods and Results 108
4.5 Classification Across Individuals - Discussion 110
4.6 Classification Using Collective Data - Methods and Results 112
4.7 Classification Using Collective Data - Discussion 119
4.8 Discussion and Conclusions 122

Chapter 5 Conclusions 124

References 128

LIST OF TABLES

Table 3.4.1.1 Distribution of Basis Function Power For The First 10 Basis Images Computed From A 10 Second of Image Sequence - Subject 1 EC 45

Table 3.4.1.2 Average Reconstruction Correlation For 10 Second Segments of Data Record - Subject 1 EC 57

Table 3.4.1.3 Average Reconstruction Correlation Corresponding to Figure 3.4.1.11 - Subject 1 EP 59

Table 3.4.3.1 Summary of Average, Maximum, and Minimum ARC Values Corresponding to Figure 3.4.3.3 70

Table 3.5.1.1 Average, Maximum, and Minimum ARC Values For Figure 3.5.1.1 77

Table 3.5.3.1 Average, Maximum, and Minimum ARC Values For Figure 3.5.3.1 82

Table 3.5.5.1 Normalized Power For the First 10 Basis Images Computed from the 50 Second Collective Data Segment 87

Table 3.5.5.2 Average, Maximum and Minimum ARC Values For Figure 3.5.5.2 87

Table 4.2.1 Classification Results For Subject 1 105

Table 4.2.2 Classification Results For Subjects 2 - 5 105

Table 4.4.1 Classification Results Using Measurement Space and Transformation Matrix of Subject 1 110

Table 4.6.1 Classification Results Using Transform Matrix Based On Collective Data Set 114

Table 4.6.2 Classification Results Using Subjects Own Training Vectors and Collective Transformation Matrix 119

LIST OF FIGURES

Figure 2.1.1.1 Linear Matrix Transformation - A Graphical View 11

Figure 2.2.1 Main Steps in Pattern Recognition 14

Figure 3.0.1 Scalp Electrode Placement 32

Figure 3.4.1.1 Mean Image Computed From First 10 Seconds of Image Sequence - Subject 1 EC 47

Figure 3.4.1.2 Basis Image 1 Computed From First 10 Seconds of Image Sequence - Subject 1 EC 48

Figure 3.4.1.3 Basis Image 2 Computed From First 10 Seconds of Image Sequence - Subject 1 EC 49

Figure 3.4.1.4 Basis Image 3 Computed From First 10 Seconds of Image Sequence - Subject 1 EC 50

Figure 3.4.1.5 Coefficient Sequence of Basis Image 1 during first 2 Seconds of Analysis - Subject 1 EC 51

Figure 3.4.1.6 Reconstruction Parameters Using First 5 Basis Images For First 2 Seconds of Subject 1 EC Data Record 52

Figure 3.4.1.7 Image Reconstruction Using the First Five Basis Images 53

Figure 3.4.1.8 Basis Image Power Representation For Varying Lengths of Data - Subject 1 EC 55

Figure 3.4.1.9 Distribution of Basis Image Power For 10 Second Segments of Data Record - Subject 1 EC 56

Figure 3.4.1.10 Power Represented by Optimal Basis Images for Varying Lengths of Data - Subject 1 EP 58

Figure 3.4.1.11 Distribution of Basis Function Power when Basis Images Computed From the First 10 Seconds are Applied To Other 10 Second Segments of Data Record - Subject 1 EP 58

Figure 3.4.3.1 Basis Images 1-3 For Subjects 2,3,4,5 EC 67

Figure 3.4.3.2 Power Represented by Optimal Basis Images For Varying Lengths of Data 68

Figure 3.4.3.3 Power Represented by Basis Images Computed From the First 10 Seconds For 10 Second Segments of the EEG Record 69

Figure 3.5.1.1 Power Represented by EC Computed Basis Images for 10 Second Segments of EP Data 76

Figure 3.5.3.1 Power Represented by Subject 1 EC Basis Images for 10 Second Segments of EC Data 81

Figure 3.5.5.1 Basis Images 1 - 5 For Collective Data Set 85

Figure 3.5.5.2 10 Second Segments of Data Projected on to Collective Basis Images 86

Figure 4.2.1 Pattern Classification for Subject 1 Using Training Data from Subject 1 104

Figure 4.6.1 Normalized Projection of EC and EP Pattern Vectors on to the First Two Principal Component Axes of the Feature Vector Set 116

Figure 4.6.2 Composite Projection of EC and EP Pattern Vectors on to the First Two Principal Component Axes of the Feature Vector Set 117

Figure 4.6.3 Normalized Projection of Transformed Data on to Principal Component Axes Used in Figures 4.6.1 and 4.6.2 118

Chapter 1

The Electroencephalogram

1.0 Introduction

Electric fields on the human scalp were first discovered by the German psychologist Hans Berger (1928). Since then much effort has been dedicated to understanding the physiological sources of these fields and correlating them to brain function. Measured using electrodes placed on the scalp, the electroencephalogram (EEG) provides a non-invasive view of the functioning brain with millisecond resolution. It is for these reasons that Hans Berger thought the EEG was the "window on the mind".

This section presents the types of brain signals typically analyzed and discusses the spatial and temporal properties of the EEG. The technique used in this thesis to investigate the EEG, the Karhunen-Loeve transform, is also briefly reviewed.

The two main types of EEG records typically analyzed are background potential and evoked (often called event related) potential data. Background EEG is taken from subjects asked to assume a particular cognitive state, and usually consists of minutes of recorded data. For example, subjects may be relaxed, and alert with their eyes closed, or may be asked to perform spatial, mathematical or language specific tasks.

Evoked potential data, on the other hand, consists of the average of many short (less than 1 second) EEG segments synchronized with a specific stimulus or response. These include flashing checkerboard patterns, or auditory tones and clicks. By averaging many trials the background EEG is eliminated and the effects of the stimulus can directly be measured. In this thesis only background EEG data is used, and therefore the term EEG will refer specifically to background EEG, unless otherwise stated.

Because the recording electrodes measure electric potential, both background and event related EEG require a voltage reference. Often an electrode on one or both ears is used to provide the reference, although neither is electrically neutral. Since the choice of reference effects the potential recordings, much effort has been directed to using reference independent measurements (Nunez, 1981).

Traditionally the recordings from each electrode on the scalp are viewed collectively on a strip chart. This method of presentation stresses the temporal nature of the signal. Consequently, most analysis favours decomposition of the EEG signals into sinusoidal components via the Fourier transform. Several frequency bands are typically used: delta (less than 4 Hz), theta (4-7 Hz), alpha (8-13 Hz), beta (14-30 Hz) and gamma (above 30 Hz).

Correlates between the power of the above frequency bands and subject state have been found. For example, the power in the alpha band is usually higher in a subject who is awake yet relaxed and has her/his eyes closed, as opposed to when her/his eyes are open. Drugs also affect the EEG. For instance, the anaesthetic halothane induces near perfect sinusoidal oscillation at several frequencies. These frequencies depend upon the blood concentration of the halothane (Nunez, 1981). In addition, epileptic patients undergoing a seizure demonstrate erratic EEG activity which manifests itself in the upper frequency bands.

Series of images representing the changing topology of potential over the scalp may be created by accounting for the potentials measured at the electrode sites and the spatial location of the electrodes. Two dimensional techniques such as bilinear and bicubic spline interpolation are typically used to reconstruct the spatial voltage distributions. Stressing the spatial properties of the EEG is becoming increasingly popular and has several advantages over conventional strip chart techniques. Two main advantages of this approach are the synthesizing of information from each channel into a "holistic" picture, and the possibility of more accurate anatomical mapping.

Furthermore, it is often possible in the spatial domain to remove the effects of an active reference electrode. For example, use of the spatial Laplacian operator not only removes dependence on a reference electrode but also provides information about current source generators thought to be responsible for much EEG activity (Nunez, 1981). The spatial maps are often transformed into two dimensional power spectrum images by applying the 2D Fourier transform or other non-linear power spectrum estimation techniques. These methods isolate the effects of an active recording reference to the (0,0) spatial wave.

Two disadvantages arise from the use of topographic maps. First, adequate spatial sampling, required to create both accurate and precise images (Koles and Paranjape, 1987), is often expensive and computationally intensive. Second, the temporal properties of the signal are neglected.

In this thesis a statistical technique known as the Karhunen-Loeve transform (KLT) is used to bridge the gap between the spatial and temporal properties of the EEG. When applied to temporal sequences of spatial EEG maps the KLT yields two products. The first, a set of orthonormal basis images, relates to the spatial nature of the EEG. These basis images represent the underlying and persistent spatial patterns in the temporal sequence of original images and are ordered in terms of their significance. That is, the first basis image accounts for the largest power possible in the original data, in a least squares sense; the second basis image the next largest amount of power, etc..

The second product of KLT analysis, a set of orthogonal coefficient sequences, indicates the amount of each basis image present in the original data images, and therefore provides the measure of significance used above. Each coefficient sequence can be thought of as extracting the temporal components common to all of the electrode recordings, and thus relates to the temporal nature of the EEG.

The mathematical development of the Karhunen-Loeve transform is presented in Chapter 2 along with an overview of pattern recognition. In Chapter 3 the KLT is used as a pattern representation technique to determine the underlying spatial properties of the EEG, while Chapter 4 uses the KLT along with another transform (Fukunaga and Koontz, 1970) as a pattern classification technique to differentiate between eyes closed and eyes open EEG. The conclusions of this research and suggestions for further investigation are presented in Chapter 5. The remainder of this chapter provides an overview of the physiological sources of the EEG.

1.1 Origin of Scalp Voltage Potentials

This section overviews the origins of the EEG. The overview is cursory since this thesis emphasizes pattern representation and classification of the EEG signals and not their physiological basis.

The brain is largely composed of nerve cells. Each nerve cell is called a neuron, and consists of a cell body, dendrites and an axon. The dendrites are typically extensions of the cell body which facilitate reception of signals from other neurons at intercellular junctions called synapses. Synaptic inputs from other neurons may either be excitatory or inhibitory. With sufficient stimulation the neuron "fires", sending an action potential (electrical charge) along its axon, which may again either stimulate or inhibit other neurons from firing.

The potential distribution around a neuron may be approximated by a dipole, that is, current flow between two poles of opposite electrical charge (Nunez, 1981). The distribution around a group of parallel and synchronized neurons may also be approximated by a dipole. In both cases the potential decreases with the inverse square of the distance from the dipole source, and is also a function of the homogeneity of the surrounding medium.

The human brain is composed of many anatomically identifiable structures, primary among which are the brainstem, cerebellum and cerebrum (Nunez, 1981). Our focus is on the cerebrum, which composes the upper portion of the brain and is divided into two hemispheres. The outer several millimeters of the cerebrum, referred to as the cerebral cortex, contains many highly connected neurons. Because of the prominence of cell bodies in the cortex it appears grey in cross section and is appropriately called grey matter. Beneath the cortex is the white matter composed of axons which connect various regions of the cortex.

Neurons in the cerebral cortex are typically pyramidal in shape and are arranged in a laminar and columnar fashion (Katznelson, 1981). The electrical activity in those cortical columns of diameter 20-50 μm is highly correlated, and it is postulated that these columns (named "minicolumns") form the basic functional units of the cortex. Because the cortex is the anatomical structure closest to the scalp and contains synchronized neural activity capable of generating observed scalp potentials, it is thought responsible for most of the EEG.

The goal of studying the EEG is determination of the underlying physiological sources of the signal and their relation to cognition and brain function. The EEG is a measure of the electric potential on the scalp, and is therefore a measure of the total electrical activity around the recording site (not a single neuron's action potential). Nunez (1981) states that the potentials measured on the scalp may originate from three sources: a single dipole, a dipole sheet, and the field from an action potential.

Many problems must be overcome to locate these dipole sources. Scalp and skull conductivity and the non-spherical nature of the skull are two effects which must be accounted for. The physiological basis of dipole current generators and the validity of assuming them as the source of the EEG also require investigation. The fundamental question, then, requires uncovering the relationship between cognitive processing and the synchronized activity of the many neurons which generate observed scalp potentials.

Although much has been accomplished in understanding the EEG, many questions still remain. Thus, it remains a corroborative clinical tool rather than one for primary diagnosis. As Gevins (1984) states, "the most important limiting factor is the lack of basic knowledge about the origin and significance of BEMS [brain electromagnetic signals]" (p. 834). The research in this thesis, while not purporting to explain the origins of the EEG, provides an additional tool to aid in decoding the significance of the EEG and its relation to the cognitive state of the subject.

Chapter 2

The Karhunen-Loeve Transform

2.0 Introduction

The Karhunen-Loeve transform (KLT), also known as principal components analysis (PCA) is the main tool for investigating the spatial EEG used in this thesis. The KLT has many desirable properties and is successfully used in multivariate statistical analysis and statistical pattern recognition. Before reviewing the mathematical details of the KLT and discussing its properties, an overview of data transformation and pattern recognition is given. Chapter 3 deals with issues specific to using the KLT on the spatial EEG.

2.1 Data Transformation

The transformation of a set of data from one measurement space to another such space is a common and useful procedure in data analysis, including statistical pattern recognition. Transformation procedures usually consist of projecting an acquired data set, taken in one measurement space, into another measurement space; in pattern recognition the latter space is typically referred to as the feature space.

In mathematical terms the transformation is written as

$$\Psi : \mathbf{X} \Rightarrow \mathbf{Y} \tag{2.1.1}$$

where : Ψ is the transformation function, **X** is the original data set, and **Y** the transformed data.

The motivation behind such a projection is to illuminate information inherent in the original data, but which is not obvious when viewing **X** in its raw form. The transformed data may provide clues as to the underlying structure of the process being measured, or provide features useful in classifying or representing the process. When **Y** is used to classify data it typically is of a lower dimension than the original data set with as much discriminatory information compressed into it as possible.

Another reason for using transformations is that mathematical operations too complex or computationally intensive to perform in the original measurement space may have a simple counterpart in the new space. Thus, if an inverse (Ψ^{-1}) to Ψ exists, the original data may be transformed into the new space using Ψ, the equivalent operation performed, and Ψ^{-1} applied to transform the processed data back to the original space. A common example involving the Fourier transform is the equivalence between multiplication in the frequency domain and convolution in the time domain.

Data transformations, or as they are sometimes called, mappings, are either linear or non-linear. In a linear mapping the transformed data set consists of linear combinations of the input data, with the transformation following the superposition principle. Non-linear mappings are in general more difficult to express mathematically and also tend to be heuristic in nature (G. Biswas et. al, 1981). The Karhunen-Loeve transform is a linear mapping and thus the rest of the chapter will concentrate on linear transformations.

2.1.1 Basic Notation of Linear Transformations

Since the analysis of data using a digital computer invariably involves working with sampled signals the following discussion is based on discrete data, represented using matrix notation. The following conventions apply: vectors are underlined, bold and in small letters; matrices are underlined, bold, and in capital letters.

Let $\underline{\mathbf{x}}_i = (x_{i,1}, x_{i,2}, \ldots x_{i,N})^t$, where superscript t (^{t}) represents the transpose operator, denote the i^{th} set of measurements, also referred to as a pattern vector, made on some random process. If there are T sets of these N-dimensional measurements acquired, a matrix $\underline{\mathbf{X}}$ can be formed, where each column of $\underline{\mathbf{X}}$ contains a different pattern vector. Each row of $\underline{\mathbf{X}}$ contains a set of numbers relating to one particular feature of the random process. Often, the T measurements represent temporal samples of some process, with the sampling rate meeting the Nyquist criterion.

In a discrete linear mapping the set of pattern vectors $\underline{\mathbf{X}}$ is represented by the linear combination of basis functions and an associated set of weighting coefficients. In matrix notation,

$$\underline{\mathbf{X}} = \underline{\mathbf{B}}\,\underline{\mathbf{K}} \qquad (2.1.1.1)$$

where $\underline{\mathbf{X}}$ is the NxT dimensional set of measurements, $\underline{\mathbf{B}}$ is an NxM dimensional set of basis functions, and $\underline{\mathbf{K}}$ is a set of MxT weighting coefficients associated with the basis functions.

Each basis function is contained in a column of $\underline{\mathbf{B}}$, while its weighting coefficients are represented in each row of $\underline{\mathbf{K}}$. In other words the j^{th} pattern vector of $\underline{\mathbf{X}}$, that is, the j^{th} column, is represented by a weighted combination of each column of $\underline{\mathbf{B}}$, where the weight for the i^{th} basis function is the value at the i^{th} row of the j^{th} column of $\underline{\mathbf{K}}$. Figure 2.1.1.1 presents a graphical representation of this process.

Note, that the interpretation of $\underline{\mathbf{X}}$, $\underline{\mathbf{B}}$ and $\underline{\mathbf{K}}$ is arbitrary. For example, the "temporal" direction in $\underline{\mathbf{X}}$ can easily be thought of as running vertically, as opposed to horizontally, and similarly the basis functions may be represented as row vectors. As well, which matrix to say contains the basis functions and which the coefficients is a matter of convenience and the properties desired for each matrix. The development presented here is aligned with the version of the KLT most useful for analyzing the spatial EEG. The notation used is typical, but by no means universal.

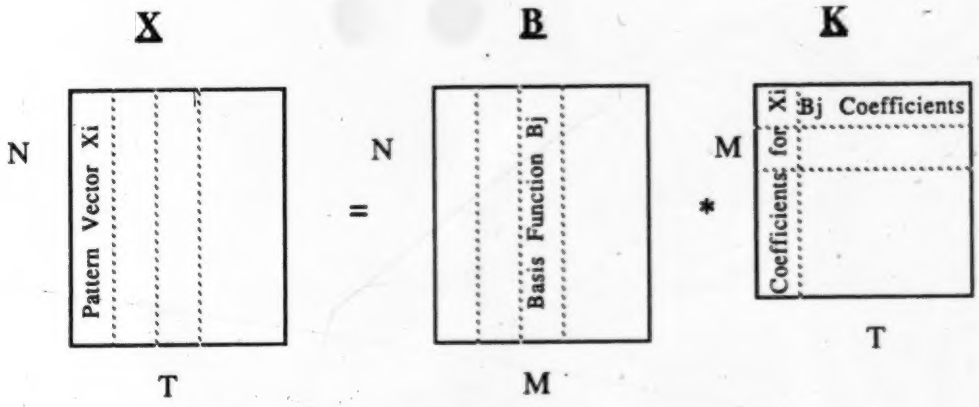


Figure 2.1.1.1
Linear Matrix Transformation - A Graphical View

In matrix notation the tth column of $\underline{\mathbf{X}}$ is written as

$$\underline{\mathbf{x}}_t = \sum_{i=1}^{M} \underline{\mathbf{b}}_i \, k_{it} \qquad (2.1.1.2)$$

while, individual elements of $\underline{\mathbf{X}}$ are calculated from the full expansion of the matrix multiplication

$$x_{n,t} = \sum_{i=1}^{M} b_{n,i} \; k_{i,t} \qquad (2.1.1.3)$$

In general, the determination of $\underline{\mathbf{B}}$ specifies the type of linear mapping. Some transformations have their basis functions determined *a priori*, that is, they are data independent, while others, including the Karhunen-Loeve transform, have process dependent basis functions. Regardless, a desirable property of the basis functions is that they represent the most information about the data set in the fewest possible elements. This property arises when the information represented by one basis function is not found in any of other basis function, that is the basis functions form an orthogonal, or orthonormal set. For an orthogonal set the dot product between each pair of basis functions is 0, while an orthonormal set has the additional constraint that the norm of each basis function is 1. Formally,

$$\underline{\mathbf{b}}_i^t \cdot \underline{\mathbf{b}}_j = \sum_k b_{ik} b_{kj} = 0 \qquad (2.1.1.4)$$

$$\underline{\mathbf{b}}_i^t \cdot \underline{\mathbf{b}}_i = \sum_k b_{ik}^2 = 1 \qquad (2.1.1.5)$$

where $\underline{\mathbf{b}}_i$ represents the ith basis function, that is, column of $\underline{\mathbf{B}}$.

One property of the linear expansion of Equation 2.1.1.1 is that if the basis functions form an orthonormal set the average power of the matrix $\underline{\mathbf{X}}$ can computed from the coefficients only. The average power is defined as

$$\mathbf{P} = \frac{1}{NT} \sum_{t=1}^{T} \underline{\mathbf{x}}_t^2 = \frac{1}{NT} \sum_{t=1}^{T} \underline{\mathbf{x}}_t^t \underline{\mathbf{x}}_t \qquad (2.1.1.6)$$

If Equation 2.1.1.2 is substituted in, and the transpose performed, then

$$\mathbf{P} = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{M} k_{it} \underline{\mathbf{b}}_i^t \sum_{j=1}^{M} \underline{\mathbf{b}}_j k_{jt} \qquad (2.1.1.7)$$

But since the basis functions are orthonormal and thus satisfy Equations 2.1.1.4 and 2.1.1.5, Equation 2.1.1.7 can be rewritten as

$$\mathbf{P} = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{M} k_{it}^2 \qquad (2.1.1.8)$$

Thus the average power in the data set matrix is calculated directly from the set of coefficients simply by squaring and averaging them over the size of the data set.

The number of coefficients required for an exact representation of the input data is $\mathbf{M}$, the value of which is determined by the basis functions used and the data itself. If a value $\mathbf{m}$, where $\mathbf{m}<\mathbf{M}$, is used instead an error, ε^2, in the reconstructed data set, $\mathbf{X}'$, occurs. If the error in representation is small it may sometimes be desirable to use $\mathbf{m}$. For example, if a tolerable error results for $\mathbf{m}<<\mathbf{M}$ then $\underline{\mathbf{X}}'$ can be stored using much less space than the original data. As well, if the data set contains noise it may not be desirable to store those basis functions and coefficients representing little variance of the original data, as they are most likely related to noise and not signal. This latter idea is important for the KLT and will be elaborated upon later.

One criterion for measuring the error in the reconstructed data set is by measuring the mean squared error (MSE), ε^2, given by

$$\varepsilon^2 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{T} (x_{i,j} - x'_{i,j})^2 \tag{2.1.1.9}$$

A common example of a linear mapping where the basis functions have been chosen *a priori*, and which is used in either discrete or continuous space, is the Fourier transform. In one dimension, the Fourier transform maps a signal on to a set of orthonormal basis functions consisting of sine and cosine waves at various frequencies. These functions are chosen because of their desirable properties in linear systems. Often, the Fourier coefficients are taken and viewed as a spectrum of the input signal. The spectrum is often used to classify the signal, for example, by the power contained in certain frequency bands, or the spectrum may be filtered by retaining those coefficients, and hence basis functions, representing a significant portion of power contained in the input signal.

2.2 Pattern Recognition

Pattern recognition is a broad field of study whose general goal is to develop methods to classify data according to which category, or class, a given sample belongs. Typically measurements are made on a sample or some process, features extracted or selected from those measurements, and a classification algorithm applied to the features to classify the data. Figure 2.2.1 illustrates the main steps performed in pattern recognition.

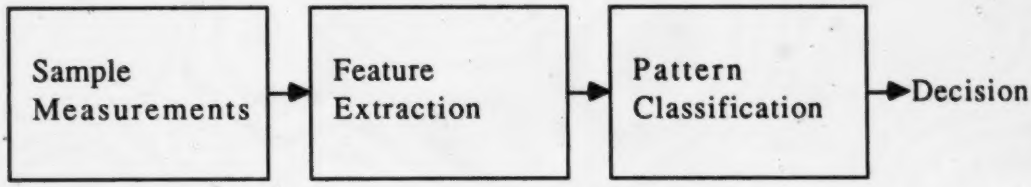


Figure 2.2.1
Main Steps in Pattern Recognition

Feature extraction can be thought of as dimensionality reduction (Kittler, 1986) since the extracted features generally number less than the measurements made on the sample. The extracted features should also have the property that they contain enough information to classify the sample according to the category of the input data. Feature extraction methods are generally broken down into two main areas: decision theoretic and structural (Hanakata, 1977). In the decision theoretic approach, the one pursued in this thesis, features basically consist of measurement vectors taken from the data set. Features may consist of selecting those measurements deemed important for classification, or of combining measurements linearly, as in the case of the Karhunen-Loeve transform, or non-linearly to produce new features used in discrimination.

In the structural approach, on the other hand, not only are measurements considered important, but also the relationships between those measurements. The features extracted tend to be high level descriptions of the object to be classified. For example, when extracting features from an image, patterns such as a line, corner, angle, or jagged edge may be found and the spatial relationships between those basic features used to categorize the sample, for example as a wrench as opposed to a screwdriver. When such descriptions are important organization of the features may proceed in a syntactic fashion, where rules of syntax are used to combine, manipulate and classify objects.

Since feature extraction in this thesis is decision theoretic, the following discussion deals with classification as it relates to decision theoretic features. It is the classifier's role to take the features extracted from the measurements and respond with the category of the data. The first basic step in classification is a "learning" phase, where samples of known class are used to "train" the classifier. In more specific terms, the parameters used by the classification method are calculated. After acquiring these parameters data of unknown class is presented to the classifier and used to measure the classification accuracy. Three basic types of classifiers exist, each differing in how much *a priori* knowledge of the feature space they assume.

When no *a priori* knowledge of the statistical distributions of the feature vectors is known deterministic techniques are used. These methods partition the feature space by various means, and assume that each feature is deterministic. A popular deterministic classification scheme is cluster analysis, where clusters of sample measurements in feature space are used to define each class. New samples are classified according various techniques, a common one being the nearest neighbour method, where an unknown sample is labeled with the same category as the training sample closest to it (using some distance measure, usually Euclidean). Linear discriminant functions are also popular, where n-dimensional hyperplanes are used to partition the (n-dimensional) feature space into feature subspaces. Unknown samples are classified according to which subspace they reside in.

The next major group of classifiers assume that the feature vectors are statistical in nature, that is, they are representative of underlying statistical processes. A major goal in this type of classification is to determine the probability density function (pdf) of each classes' measurements or extracted features. By knowing the probability of occurrence of each class along with its pdf, the classifier is designed to minimize misclassification. If these parameters are known optimum classification occurs using Bayesian estimation (Fukunaga, 1986). The obvious difficulty with this approach is that knowledge of the underlying statistical properties of each sample process is difficult to obtain. Although this is the case, estimation of the pdf's of the sample processes are usually computed in order to give bounds to the minimum misclassification error (Bayesian error). If this estimation is unacceptably high new features or measurements should be taken.

The last group of classification algorithms are the trainable pattern classifiers, where the classifier "learns" the information it requires. These classifiers may work by either using statistical or deterministic techniques, or a combination of both. Trainable classifiers may be further categorized into supervised or unsupervised classifiers. In the former class information about each training sample is known and can be used to aid in "learning", while in the latter class information is not known. In unsupervised training assumptions about the feature set distribution, for example that features cluster together (Shimura, 1973) are used to train the classifier.

2.3 Karhunen-Loeve Transform Theory

The Karhunen-Loeve transform is a linear expansion, and when applied to an **NxT** matrix $\underline{\mathbf{X}}$ yields a set of orthonormal basis functions, and in which the error in approximating the original data set by using only the first **m** basis functions is a minimum in a least squares sense. Because of the latter property the basis functions are data dependent, and hence cannot be determined *a priori*, as in the Fourier transform, for example. This property is useful, though, since the basis functions may isolate components underlying the statistical process being investigated. As well, information about the original data set is compressed onto as few basis vectors in the transformed space as possible (in a least squares sense). Thus if only a few basis functions and associated coefficients are required to represent a significant portion of the original data, that is, the mean squared error is small, then data compression may be performed by storing only the required basis functions and coefficients.

A drawback of this transform is that because the basis functions are data dependent they will most likely vary between different data sets. Thus, in comparing and classifying two or more sets of measurements not only must the coefficients be compared, as with the Fourier transform spectra, for example, but also the basis functions. This problem is addressed in Chapter 3.

2.3.1 Review of Eigen Analysis

Eigenvalues and eigenvectors play important roles in much of science and engineering. A brief review of their properties is given here, although the discussion is necessarily cursory.

A matrix $\underline{\mathbf{A}}$ is said to be a linear transformation from a vector space $\underline{\mathbf{U}}$ to a vector space $\underline{\mathbf{V}}$ if the transformation follows the properties of superposition, and if the transformed vector ($\underline{\mathbf{v}} = \underline{\mathbf{A}}\,\underline{\mathbf{u}}$) is unambiguously defined. The special class of vectors which after transformation by a square matrix $\underline{\mathbf{A}}$ remain in the vector space $\underline{\mathbf{U}}$, and which change only by a scalar multiple λ, are called the eigenvectors of $\underline{\mathbf{A}}$. The scalar multiple associated with each eigenvector is called its corresponding eigenvalue. In mathematical terms

$$\underline{\mathbf{A}}\,\underline{\mathbf{u}} = \underline{\mathbf{u}}\,\lambda \qquad (2.3.1.1)$$

Often the eigenvalues are called the characteristic values, or "proper" values of $\underline{\mathbf{A}}$.

It can easily be shown that if $\underline{\mathbf{A}}$ is real and symmetric the eigenvalues of $\underline{\mathbf{A}}$ are real, and its eigenvectors form an orthonormal set. Also, if $\underline{\mathbf{A}}$ is of size NxN then it can also be shown that there will be at most N distinct eigenvectors. Thus, if the matrix $\underline{\mathbf{U}}$ is created, where each column contains a different eigenvector, and a diagonal matrix $\underline{\boldsymbol{\lambda}}$ formed where i^{th} diagonal element corresponds to the i^{th} eigenvector (i.e. column) of $\underline{\mathbf{U}}$, Equation 2.3.1.1 can be rewritten as

$$\underline{\mathbf{A}}\,\underline{\mathbf{U}} = \underline{\mathbf{U}}\,\underline{\boldsymbol{\lambda}} \qquad (2.3.1.2)$$

or, by using the orthonormality of the eigenvectors for a real and symmetric matrix, that is, $\underline{\mathbf{U}}^{-1} = \underline{\mathbf{U}}^{t}$,

$$\underline{\mathbf{A}} = \underline{\mathbf{U}}\,\underline{\boldsymbol{\lambda}}\,\underline{\mathbf{U}}^{t} \qquad (2.3.1.3)$$

This last form is important in the development of the KLT.

2.3.2 Autocorrelation Functions

Autocorrelation functions also play a key role in the development of the KLT and will be given a brief review here. An autocorrelation function is measure of the self-similarity of a stochastic process. If **X** represents a continuous stochastic process, the similarity $\phi(ti,tj)$ between $\mathbf{X}_{ti}$ and $\mathbf{X}_{tj}$ (the i^{th} and j^{th} time instants of **X**) is measured by the expected value of the process at $\mathbf{X}_{ti}$ and $\mathbf{X}_{tj}$,

$$E\{\mathbf{X}_{ti}\,\mathbf{X}_{tj}\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_{ti}\,x_{tj}\,p(x_{ti}\,x_{tj})\,dx_{ti}\,dx_{tj} \qquad (2.3.2.1)$$

where $p(\mathbf{x}_{ti},\mathbf{x}_{tj})$ is the joint probability density function of the process and E{} the expectation operator.

Thus the autocorrelation function of a one dimensional process is a two dimensional function $\phi(ti,tj)$. If the process **X** is stationary, that is, its statistics do not change over time, then $\phi(ti,tj)$ is a function of only the difference between ti and tj, τ, not their absolute values. The autocorrelation function can therefore be written as $\phi(\tau)$.

For the discrete case the process can be represented by a matrix $\underline{\mathbf{X}}$, where each row represents a different measurement of the process, and each column ($\underline{\mathbf{X}}_t$) a different realization of the process at various values of t. The autocorrelation function in such a case is written as

$$\underline{\mathbf{R}} = \underset{t}{\mathbf{E}}\left\{\underline{\mathbf{X}}_t\,\underline{\mathbf{X}}_t^t\right\} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=-T}^{T}\underline{\mathbf{X}}_t\,\underline{\mathbf{X}}_t^t = \lim_{T\to\infty}\frac{1}{T}\underline{\mathbf{X}}\,\underline{\mathbf{X}}^t \qquad (2.3.2.2)$$

This assumes that the process being measured is stationary through time (that is, the statistics of each measurement [row] of $\underline{\mathbf{X}}$ does not vary with t), and that the *a priori* probability of occurrence of each pattern vector is equal.

If all of the process measurements (rows) have the same statistics then $\underline{\mathbf{R}}$ is Toeplitz, that is, its diagonals contain constant values.

In this case, as with the continuous process, the function may be reduced to a one dimensional matrix, since each row of $\underline{\mathbf{R}}$ is just a shifted version of a vector, $\underline{\phi}_{\mathbf{R}}$.

It can be shown that the average power in $\underline{\mathbf{X}}$ may be computed from the sum of squares of the main diagonal elements (i.e. the trace) of $\underline{\mathbf{R}}$. In other words,

$$P = \lim_{T \to \infty} \frac{1}{T} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{T} x_{ij}^{2} = \frac{1}{N} \sum_{i=1}^{N} r_{ii}^{2} \qquad (2.3.2.3)$$

2.3.3 Mathematical Development of the Karhunen-Loeve Transform

Recall Equation 2.1.1.1 which states that a NxT matrix $\underline{\mathbf{X}}$ may be represented in terms of a set of basis functions (or vectors), $\underline{\mathbf{B}}$, and an associated set of coefficients, $\underline{\mathbf{K}}$. For the KLT the expected mean squared error, ε^2, in $\underline{\mathbf{X}}'$ is a minimum. Recall that $\underline{\mathbf{X}}'$ is the reconstructed version of $\underline{\mathbf{X}}$ using **m** basis functions, where **m** is less than the number required, **M**, for exact reconstruction.

The following development assumes that the process being investigated is ergodic and therefore that in the limit **T** approaches infinity, or that it is sufficiently large so that $\underline{\mathbf{X}}$ fully captures the processes' second order statistics.

The t^{th} reconstructed column of $\underline{\mathbf{X}}'$ is written as

$$\underline{\mathbf{x}}_t' = \sum_{i=1}^{m} k_{it} \, \underline{\mathbf{b}}_i \qquad (2.3.3.1)$$

The error in reconstruction of $\underline{\mathbf{x}}_t$ is thus

$$\underline{\mathbf{x}}_t - \underline{\mathbf{x}}_t' = \sum_{i=m+1}^{M} k_{it} \, \underline{\mathbf{b}}_i \qquad (2.3.3.2)$$

The mean squared error, ε^2, of the approximation for the entire reconstructed data set is

$$\varepsilon^2 = \underset{t}{\mathbf{E}}\left\{ (\mathbf{x}_t - \mathbf{x}_t')^t\,(\mathbf{x}_t - \mathbf{x}_t') \right\} \qquad (2.3.3.3)$$

which can be rewritten using Equation 2.3.3.2 as

$$\varepsilon^2 = \underset{t}{\mathbf{E}}\left\{ \Big(\sum_{i=m+1}^{M} k_{it}\,\mathbf{b}_i\Big)^t \Big(\sum_{j=m+1}^{M} k_{jt}\,\mathbf{b}_j\Big) \right\} \qquad (2.3.3.4)$$

Since the basis functions are orthonormal they are constrained as follows

$$\mathbf{b}_i^t \cdot \mathbf{b}_j = \delta_{ij} \qquad (2.3.3.5)$$

where δ_{ij} is the Kronecker delta function

$$\delta_{ij} = \begin{cases} 1, & i=j \\ 0, & i\neq j \end{cases} \qquad (2.3.3.6)$$

Thus, Equation 2.3.3.4 can be written

$$\varepsilon^2 = \underset{t}{\mathbf{E}}\left\{ \sum_{i=m+1}^{M} k_{it}^2 \right\} \qquad (2.3.3.7)$$

The mean squared error in using **m** basis functions is therefore the expected value of the square of the coefficients of the unused basis functions. But, it is desired to express the mean squared error in terms of the data set vectors, $\mathbf{x}_t$. If both sides of Equation 2.1.1.2 are multiplied by $\mathbf{b}_j^t$ then

$$\mathbf{b}_j^t\,\mathbf{x}_t = \mathbf{b}_j^t \sum_{i=1}^{M} \mathbf{b}_i\, k_{it} \qquad (2.3.3.8)$$

and again using the orthonormality constraint of Equation 2.3.3.5

$$\mathbf{b}_j^t\,\mathbf{x}_t = k_{jt} \qquad (2.3.3.9)$$

Substituting this result into Equation 2.3.3.7 gives the mean squared error in terms of the original measurement set

$$\varepsilon^2 = \underset{t}{\mathbf{E}}\left\{ \sum_{i=m+1}^{M} (\mathbf{b}_i^t\,\mathbf{x}_t)\,(\mathbf{b}_i^t\,\mathbf{x}_t)^t \right\} \qquad (2.3.3.10)$$

$$= \underset{t}{E}\left\{\sum_{i=m+1}^{M} \underline{\mathbf{b}}_i^t \,\underline{\mathbf{x}}_t \,\underline{\mathbf{x}}_t^t \,\underline{\mathbf{b}}_i\right\} \tag{2.3.3.10a}$$

Since the basis functions are independent of time (t), the expectation can be brought inside and ε^2 written as

$$\varepsilon^2 = \sum_{i=m+1}^{M} \underline{\mathbf{b}}_i^t \,\underset{t}{E}\left\{\underline{\mathbf{x}}_t \,\underline{\mathbf{x}}_t^t\right\} \underline{\mathbf{b}}_i \tag{2.3.3.11}$$

Now note that the expectation operator when taken over t defines the autocorrelation function $\underline{\mathbf{R}}$, of $\underline{\mathbf{X}}$ (Equation 2.3.2.2). Therefore the mean squared error in reconstruction is related to the autocorrelation function of $\underline{\mathbf{X}}$ in the following way

$$\varepsilon^2 = \sum_{i=m+1}^{M} \underline{\mathbf{b}}_i^t \,\underline{\mathbf{R}}\, \underline{\mathbf{b}}_i \tag{2.3.3.12}$$

It is desired to minimize ε^2 while maintaining the orthonormality constraint of the basis functions. To solve for the $\underline{\mathbf{b}}_i$'s the method of Lagrange multipliers is used.

First the constraint is rewritten as

$$\underline{\mathbf{b}}_i^t \underline{\mathbf{b}}_i - 1 = 0 \tag{2.3.3.13}$$

Multiplying both sides of the above equation by a scalar constant λ_i does not change the equality and thus

$$\lambda_i\,(\underline{\mathbf{b}}_i^t \underline{\mathbf{b}}_i - 1) = 0 \tag{2.3.3.14}$$

or, it can equally be stated that

$$\sum_{i=m+1}^{M} \lambda_i\,(\underline{\mathbf{b}}_i^t \underline{\mathbf{b}}_i - 1) = 0 \tag{2.3.3.15}$$

The next step is to create a function $g(\underline{\mathbf{b}}_i)$ consisting of the function in which the extremum is to be found and its constraints

$$g(\underline{\mathbf{b}}_i) = \sum_{i=m+1}^{M} \underline{\mathbf{b}}_i^t \,\underline{\mathbf{R}}\, \underline{\mathbf{b}}_i \;-\; \sum_{i=m+1}^{M} \lambda_i\,(\underline{\mathbf{b}}_i^t \underline{\mathbf{b}}_i - 1) \tag{2.3.3.16}$$

To find the extremum of $g(\underline{\mathbf{b}}_i)$, and hence the values which minimize ε, $g(\underline{\mathbf{b}}_i)$ is differentiated with respect to $\underline{\mathbf{b}}_i$ and the result set to zero. In order that the mean squared error be zero $\mathbf{m}$ is set to $\mathbf{M}$, and thus there are $\mathbf{M}$ basis functions ($\underline{\mathbf{b}}_i$'s) and hence derivatives to compute. Since each $\underline{\mathbf{b}}_i$ is an independent variable Equation 2.3.3.16 may be simplified for each basis function to

$$g(\underline{\mathbf{b}}_i) = \underline{\mathbf{b}}_i^t \underline{\mathbf{R}}\, \underline{\mathbf{b}}_i - \lambda_i (\underline{\mathbf{b}}_i^t \underline{\mathbf{b}}_i - 1) \qquad (2.3.3.17)$$

Performing the differentiation with respect to $\underline{\mathbf{b}}_i$ yields

$$\underline{\mathbf{b}}_i^t \underline{\mathbf{R}} - \lambda_i \underline{\mathbf{b}}_i^t = 0 \qquad (2.3.3.18)$$

Rewriting for all $\mathbf{M}$ basis functions and corresponding eigenvalues, as was done in Section 2.3.1, and recalling the orthonormality of the basis functions ($\underline{\mathbf{B}}\, \underline{\mathbf{B}}^t = \underline{\mathbf{B}}^t \underline{\mathbf{B}} = \underline{\mathbf{I}}$) yields

$$\underline{\mathbf{B}}^t \underline{\mathbf{R}} = \underline{\boldsymbol{\lambda}}\, \underline{\mathbf{B}}^t \qquad (2.3.3.19)$$

$$\underline{\mathbf{R}} = \underline{\mathbf{B}}\, \underline{\boldsymbol{\lambda}}\, \underline{\mathbf{B}}^t \qquad (2.3.3.20)$$

By comparing Equation 2.3.3.20 and 2.3.1.3 it is obvious that each basis function is an eigenvector of the autocorrelation function $\underline{\mathbf{R}}$. However, this result does not show which eigenvalues and eigenvectors of $\underline{\mathbf{R}}$ are needed to represent $\underline{\mathbf{X}}$ with the minimum mean squared error, only that they form the set which does so.

If Equation 2.3.3.20 is rewritten for each basis function

$$\underline{\mathbf{R}}\, \underline{\mathbf{b}}_i = \lambda_i \underline{\mathbf{b}}_i \qquad (2.3.3.21)$$

and substituted into Equation 2.3.3.12, ε^2 can be written as

$$\varepsilon^2 = \sum_{i=m+1}^{M} \underline{\mathbf{b}}_i^t \lambda_i \underline{\mathbf{b}}_i \qquad (2.3.3.22)$$

Since λ_i is a constant and the $\underline{\mathbf{b}}_i$'s are orthonormal the mean squared error in reconstruction is

$$\varepsilon^2 = \sum_{i=m+1}^{M} \lambda_i \qquad (2.3.3.23)$$

Thus the mean squared error in not using eigenvector $\underline{\mathbf{b}}_i$ is its corresponding eigenvalue. If the eigenvalues, and hence eigenvectors are ordered such that

$$\lambda_1 \geq \lambda_2 \geq \ldots\ldots \geq \lambda_M \qquad (2.3.3.24)$$

selecting the first **m** eigenvectors results in the smallest error.

In order to calculate the coefficients for the linear expansion the orthonormal property of the basis functions is used to yield

$$\underline{\mathbf{K}} = \underline{\mathbf{B}}^{-1}\,\underline{\mathbf{X}} = \underline{\mathbf{B}}^{t}\,\underline{\mathbf{X}} \qquad (2.3.3.25)$$

To compute the coefficients therefore amounts to projecting the data set onto the set of axes defined by $\underline{\mathbf{B}}$. Viewed in this way the Karhunen-Loeve transform is a method to project a data set onto an optimized coordinate system, where as much information (power) of the data set is represented on as few axes as possible.

The relationship between the coefficients and eigenvalues is found by expanding the expectation operator in Equation 2.3.3.7 (assuming that all coefficients occur with equal *a priori* probability of occurrence) and summing over all basis functions

$$\varepsilon^2 = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{i=m+1}^{M} k_{it}^2 \qquad (2.3.3.26)$$

Comparing this with Equation 2.3.3.23 it is easy to see the eigenvalues are related to the coefficients by

$$\lambda_i = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} k_{it}^2 \qquad (2.3.3.27)$$

In other words, the i^{th} eigenvalue of $\underline{\mathbf{R}}$ is the mean squared value of the coefficient sequence for the i^{th} basis function. Moreover, since the basis functions each have unit power λ_i represents the power contributed by the i^{th} basis function's coefficient sequence.

A further important result is found by summing over all eigenvalues and averaging over **N**,

$$\frac{1}{N}\sum_{i=1}^{M}\lambda_i = \lim_{T\to\infty}\frac{1}{T}\frac{1}{N}\sum_{i=1}^{M}\sum_{t=1}^{T}k_{it}^2 \qquad (2.3.3.28)$$

This result is identical to the average power in a matrix $\underline{\mathbf{X}}$ for an orthonormal linear expansion (Equation 2.1.1.8), again assuming that $\underline{\mathbf{X}}$ is infinite length, or of sufficient length to capture the second order statistics of the process being measured. These last two properties of the KLT are important and used extensively in pattern recognition and representation of the EEG in Chapter 3.

To summarize the section in brief then,

- the Karhunen-Loeve basis functions are the eigenvectors of the autocorrelation function, $\underline{\mathbf{R}}$, of $\underline{\mathbf{X}}$,
- the coefficients for those basis functions are computed by projecting the original data set into the space defined by those eigenvectors,
- the mean squared error in using only the first **m** basis functions is found by summing the eigenvalues of the unused basis functions (assuming proper ordering of the eigenvectors),
- the i^{th} eigenvector of $\underline{\mathbf{R}}$ gives the power represented by the i^{th} basis function, while summing all eigenvalues gives the total power in $\underline{\mathbf{X}}$.

With these basic properties in mind the next section deals with some practical considerations in computing the basis functions and coefficients, and gives further properties of the expansion.

2.3.4 Additional Properties and Remarks

The previous section described the Karhunen-Loeve transform assuming that the exact autocorrelation matrix, $\underline{\mathbf{R}}$, of the process could be measured. When dealing with finite data lengths, or processes which are not stationary over all time, the KLT may still be applied. In this case though, the matrix **R** will represent the autocorrelation matrix of the data matrix **X**, and only an estimate of true autocorrelation matrix of the process being investigated. The products of the KLT will therefore be data dependent rather than process dependent. In other words, the basis functions will span the sample data, and not necessarily the process, with the minimum least squared error.

For a finite data set the autocorrelation matrix may be written in one of two ways, either by correlating through time (**R**) or process measurement (**Z**). Each of these matrices contain the same information, although it is distributed differently in each. Assuming equal *a priori* probabilities of occurrence for each pattern vector $\mathbf{x}_i$, **R** and **Z** may be computed from either of the following forms

$$\mathbf{R} = \frac{1}{T}\mathbf{X}\mathbf{X}^t \qquad (2.3.4.1)$$

$$\mathbf{Z} = \frac{1}{N}\mathbf{X}^t\mathbf{X} \qquad (2.3.4.2)$$

Of prime importance is the determination of M, the number of basis functions required to exactly represent the data set **X**. This depends upon the correlations inherent within **X**, more specifically upon the rank of the matrix. The rank of a matrix is a measure of the number of dimensions, that is linearly independent rows or columns, contained in the matrix. To represent a matrix using a set of basis functions (vectors) requires no more basis functions than the rank of the matrix. For an AxB matrix the rank of the matrix is, at most, the smaller of A or B.

Thus, **X**, of size NxT, requires no more than the smaller of N or T basis functions. To compute the eigenvectors of **X** it was shown in Section 2.3.3 required eigen analysis of **R**, which is of size NxN. If the rank of **X** is T, then (N-T) of the eigenvalues and eigenvectors of **R** will be zero and computational inefficiencies will result, and may be severe if **N** is much larger than **T**.

To avoid this problem recall that the autocorrelation function may also be expressed as $\underline{\mathbf{Z}}$, of size TxT, and contains the same information about $\underline{\mathbf{X}}$ as $\underline{\mathbf{R}}$. By performing eigenanalysis on $\underline{\mathbf{Z}}$ it should also be possible to compute the same basis functions as from $\underline{\mathbf{R}}$. To see how this is possible the relationships between the basis functions, coefficients and autocorrelation matrices is explored.

To get $\underline{\mathbf{R}}$ in terms of the basis functions and coefficients the linear expansion of $\underline{\mathbf{X}}$ from Equation 2.1.1.1 is substituted into the definition of $\underline{\mathbf{R}}$ from Equation 2.3.2.2 giving

$$\underline{\mathbf{R}} = \frac{1}{T}(\underline{\mathbf{B}}\,\underline{\mathbf{K}})(\underline{\mathbf{B}}\,\underline{\mathbf{K}})^{t} = \frac{1}{T}\underline{\mathbf{B}}\,\underline{\mathbf{K}}\,\underline{\mathbf{K}}^{t}\underline{\mathbf{B}}^{t} \qquad (2.3.4.3)$$

Comparing this with Equation 2.3.3.20 note that if the coefficient matrices are grouped along with the scalar constant then

$$\underline{\boldsymbol{\lambda}} = \frac{1}{T}\underline{\mathbf{K}}\,\underline{\mathbf{K}}^{t} \qquad (2.3.4.4)$$

which is just the matrix form of Equation 2.3.3.27.

Recall that $\underline{\boldsymbol{\lambda}}$ is a diagonal matrix, and thus the coefficient matrix of the KLT is an orthogonal set and can be made orthonormal by appropriate normalization. In effect then, the role of $\underline{\mathbf{B}}$ and $\underline{\mathbf{K}}$ can be interchanged by normalizing $\underline{\mathbf{K}}$ and shifting the normalizing factor into $\underline{\mathbf{B}}$. Depending upon the interpretation of the signal space $\underline{\mathbf{X}}$ desired, this view, or the one derived earlier can be used. This is a powerful property of the KLT, and will be discussed in relation to the spatial and temporal EEG in Chapter 3.

Now, to see how $\underline{\mathbf{B}}$ and $\underline{\mathbf{K}}$ relate to $\underline{\mathbf{Z}}$, Equation 2.1.1.1 is substituted into the definition of $\underline{\mathbf{Z}}$ to give

$$\underline{\mathbf{Z}} = \frac{1}{N}(\underline{\mathbf{B}}\,\underline{\mathbf{K}})^{t}(\underline{\mathbf{B}}\,\underline{\mathbf{K}}) = \frac{1}{N}\underline{\mathbf{K}}^{t}\underline{\mathbf{B}}^{t}\underline{\mathbf{B}}\,\underline{\mathbf{K}} \qquad (2.3.4.5)$$

Using the orthogonality constraint on $\underline{\mathbf{B}}$ Equation 2.3.4.5 can be simplified to

$$\underline{\mathbf{Z}} = \frac{1}{N}\underline{\mathbf{K}}^t\underline{\mathbf{K}} = \left(\frac{1}{\sqrt{N}}\underline{\mathbf{K}}^t\right)\left(\frac{1}{\sqrt{N}}\underline{\mathbf{K}}\right) \tag{2.3.4.6}$$

Performing eigenanalysis on $\underline{\mathbf{Z}}$ yields (using different variables to differentiate the eigenvalues and eigenvectors from those of $\underline{\mathbf{R}}$)

$$\underline{\mathbf{Z}} = \underline{\mathbf{U}}\,\underline{\boldsymbol{\lambda}}_z\,\underline{\mathbf{U}}^t \tag{2.3.4.7}$$

Since $\underline{\mathbf{Z}}$ is real and symmetric the eigenvectors $\underline{\mathbf{U}}$ are orthonormal and $\underline{\boldsymbol{\lambda}}_z$ is a diagonal matrix. Thus a new matrix $\sqrt{\underline{\boldsymbol{\lambda}}_z}$ can be created in which the ith diagonal element is the square root of ith diagonal element of $\underline{\boldsymbol{\lambda}}_z$. It is therefore possible to write

$$\underline{\boldsymbol{\lambda}}_z = \sqrt{\underline{\boldsymbol{\lambda}}_z}\,\sqrt{\underline{\boldsymbol{\lambda}}_z^t} \tag{2.3.4.8}$$

Using this relation in Equation 2.3.4.7 gives

$$\underline{\mathbf{Z}} = \left(\underline{\mathbf{U}}\sqrt{\underline{\boldsymbol{\lambda}}_z}\right)\left(\sqrt{\underline{\boldsymbol{\lambda}}_z^t}\,\underline{\mathbf{U}}^t\right) = \left(\underline{\mathbf{U}}\sqrt{\underline{\boldsymbol{\lambda}}_z}\right)\left(\underline{\mathbf{U}}\sqrt{\underline{\boldsymbol{\lambda}}_z}\right)^t \tag{2.3.4.9}$$

Comparing each bracketed term above with Equation 2.3.4.6 and solving for $\underline{\mathbf{K}}$ yields

$$\underline{\mathbf{K}} = \sqrt{N}\sqrt{\underline{\boldsymbol{\lambda}}_z^t}\,\underline{\mathbf{U}}^t \tag{2.3.4.10}$$

Thus, the coefficient sequence can be computed from the autocorrelation function $\underline{\mathbf{Z}}$, as well as $\underline{\mathbf{R}}$. In the latter case the basis functions can be solved by rewriting Equation 2.3.1.1 giving $\underline{\mathbf{B}} = \underline{\mathbf{X}}\,\underline{\mathbf{K}}^{-1}$. To minimize computational costs, though, it is desirable to avoid this matrix inversion.

Using relationship 2.3.4.4 and post multiplying both sides of Equation 2.1.1.1 by $\underline{\mathbf{K}}^t$ relates $\underline{\mathbf{B}}$ to $\underline{\boldsymbol{\lambda}}$ and $\underline{\mathbf{K}}^t$ in the following manner

$$\underline{\mathbf{B}}\,\underline{\boldsymbol{\lambda}} = \frac{1}{T}\underline{\mathbf{X}}\,\underline{\mathbf{K}}^t \tag{2.3.4.11}$$

Since $\underline{\boldsymbol{\lambda}}$ is a diagonal matrix $\underline{\boldsymbol{\lambda}}^{-1}$ is a diagonal matrix with the ith diagonal element equal to $1/\lambda_i$ (which involves no matrix inversion). If $\mathbf{\lambda}^{-1}$ is denoted as $1/\mathbf{\lambda}$ then Equation 2.3.4.11 can be rewritten as

$$\mathbf{B} = \frac{1}{T}\mathbf{X}\mathbf{K}^t\frac{1}{\mathbf{\lambda}} \qquad (2.3.4.12)$$

The last remaining step in therefore to show the relationship between $\mathbf{\lambda}_z$ and $\mathbf{\lambda}$. Substituting Equation 2.3.4.10 into 2.3.4.4 gives

$$\begin{aligned}\mathbf{\lambda} &= \frac{1}{T}\left(\sqrt{N}\sqrt{\mathbf{\lambda}_z^t}\,\mathbf{U}^t\right)\left(\sqrt{N}\sqrt{\mathbf{\lambda}_z^t}\,\mathbf{U}^t\right)^t \\ &= \frac{N}{T}\sqrt{\mathbf{\lambda}_z^t}\,\mathbf{U}^t\,\mathbf{U}\sqrt{\mathbf{\lambda}_z} \qquad (2.3.4.13)\end{aligned}$$

Recalling the orthonormality of the eigenvectors of a real and symmetric matrix (which, of course, $\mathbf{Z}$ is) and using Equation 2.3.4.8 the following simple relationship emerges

$$\mathbf{\lambda} = \frac{N}{T}\mathbf{\lambda}_z \qquad (2.3.4.14)$$

Combining 2.3.4.12 and 2.3.4.14 gives the desired equation

$$\mathbf{B} = \frac{1}{N}\mathbf{X}\mathbf{K}^t\frac{1}{\mathbf{\lambda}_z} \qquad (2.3.4.15)$$

In summary then, using Equations 2.3.4.10 and 2.3.4.15 after performing eigen analysis on $\mathbf{Z}$ gives the same basis functions and coefficients as using $\mathbf{R}$. The advantage is that $\mathbf{Z}$ provides a more computationally efficient means for computing $\mathbf{B}$ and $\mathbf{K}$ when T (the size of $\mathbf{Z}$) is smaller than N (the size of $\mathbf{R}$).

The other important point is that the coefficient matrix derived using the Karhunen-Loeve transform is orthogonal, and thus by appropriate normalization either $\mathbf{B}$ or $\mathbf{K}$ may take the role of the basis function matrix or coefficient matrix.

Chapter 3

The Karhunen-Loeve Transform Applied to Pattern Representation of The Spatial EEG

3.0 Introduction

In this thesis the Karhunen-Loeve transform (KLT), also known as Principal Components Analysis (PCA), is applied to the problem of analyzing temporal sequences of the spatial EEG. In statistical pattern recognition (Fukunaga, 1972) the KLT is used to extract features to both represent and classify data. This chapter concentrates on using the KLT to represent temporal sequences of spatial EEG maps (pattern representation), while Chapter 4 uses the KLT to extract features for classification of the maps (pattern classification).

Unlike other studies which concentrate on applying the technique to evoked potential data (Skrandies and Lehmann, 1982; Kavanagh et al. 1976; Donchin and Heffley 1978), or to maps corresponding to various temporal frequencies (Nunez 1981), the KLT is applied here to unfiltered background spatial EEG recorded under various cognitive states. Towards this end, three main questions have been addressed, each building upon the previous, namely:

1) Can the Karhunen-Loeve transform be used to represent the spatial EEG in a lower dimensional space than the measurement space (i.e. does the spatial EEG have an intrinsic spatial dimensionality)? The results of this yield basis images and coefficients which represent the temporal image sequences.

2) Are the modes of the spatial EEG (manifested in the basis functions and coefficients, found above) common or different among various individuals and between cognitive states?

3) Can the basis functions and/or coefficients be used to discriminate between various cognitive states?

The first of these three questions asks in how many basis functions (images) is the spatial EEG adequately represented. These basis images may represent underlying physiological sources of the EEG, such as current generators, however it is not the purpose of this research to investigate this area. The second question extends the first by asking if the basis functions computed above are shared by cognitive states and individuals. Recall from Chapter 2 that the basis functions are unique to every data set analyzed. Therefore a problem results in comparing images and coefficient sequences. This issue is discussed in detail in Section 3.4.

The third, and most difficult question to answer, asks if the basis functions and coefficients can be used to classify the data according to cognitive state. The method used is based upon the theory presented by Fukunaga and Koontz (1970).

In this chapter the results of using the KLT for pattern representation of the spatial EEG is presented, with the results of using the KLT for pattern classification presented in Chapter 4. Before giving the results and discussion of this research the problem of applying the KLT to image sequences is covered as well as a description of the data acquisition process, and notes about the programs used to analyze the data. The notation used is the same as that in Chapter 2, except where noted.

3.1 Data Acquisition and Experimental Set Up

The data used to investigate the questions posed in 3.0 was obtained at the University of Alberta Hospital's EEG laboratory. EEG data were recorded from healthy fe/male volunteers using 31 electrodes (excluding a single ear reference) placed in the non-corner points of a 5x7 rectangular grid over the scalp. Figure 3.0.1 provides a graphical view of this sampling montage.

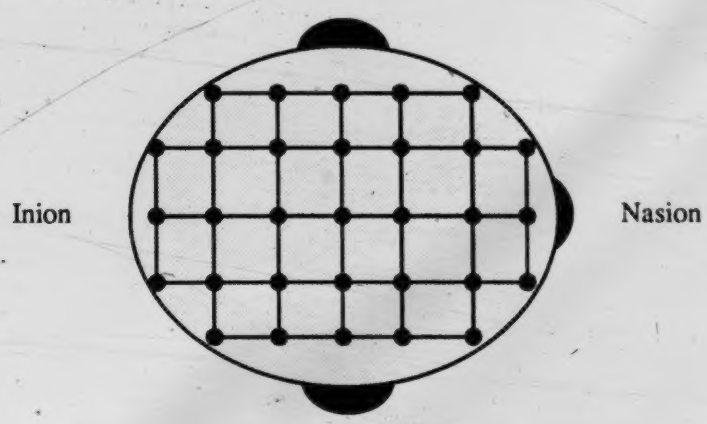


Figure 3.0.1
Scalp Electrode Placement

The data was sampled at 120 cps, digitized to 12 bits and transferred to a VAX/11-750 computer for analysis. Only artifact free portions of the recorded data were used in the subsequent analysis. Temporal filtering to various frequency bands, e.g. alpha, beta, e.t.c, was not performed before analysis, although high pass filtering was used to make each channel recording zero mean. After high pass filtering each sampled set of data corresponding to the above montage defined what is heretofore termed a "raw" image.

Data from each subject was obtained in 4 cognitive states, with approximately 2 minutes of EEG recorded per state. In the first state, denoted EC, the subjects eyes were closed, and they were asked to adopt a relaxed, resting mental state, though they were to remain alert. The second state, EP, was the same as EC except that the subjects eyes were open and they were told to visually fixate on a single spot. In the third state, LG, the subjects were asked to perform language comprehension tasks, while in the fourth state, GE, subjects were asked to perform spatial rotation problems.

Data from the latter two cognitive states are not used in this thesis for two main reasons. First, much of the LG and GE data contains artifacts and would have to be carefully screened before analysis, with the chances that lengthy, that is 5 seconds or more, stretches of continuous data would be difficult to find. The second reason that classification and analysis is performed on the EC and EP states is because these the two states are generally easily classifiable by clinicians. Thus the performance of the methods used here can be compared to that of trained clinicians.

3.2 Analyzing Image Sequences with the KLT

Recall from Chapter 2 that the KLT can be used to analyze a set of N measurements made on a process, where each measurement is considered a temporally sampled signal (where there are, say, T samples made). The difficulty in using this approach (for **T** images) is that there are no longer N signals where order is unimportant, but rather a **nxm** set of signals (i.e. pixels) in which spatial relations between pixels is important. In other words, the problem amounts to performing a 3 dimensional transform (two dimensions for each image, plus one temporal dimension for the set of images).

In order to use the relations discussed in Chapter 2, where the data set was two dimensional, the three dimensional problem in analyzing image sequences must be reduced into a two dimensional one. The method used here is drawn from Rosenfeld and Kak (1982).

Recall from Chapter 2 that each basis function is a column of matrix $\underline{\mathbf{B}}$, where each column is of size N. Thus, any relations between the rows of $\underline{\mathbf{X}}$, the data matrix, are preserved, though transformed, in the rows of $\underline{\mathbf{B}}$. It is easy to see then if each column of $\underline{\mathbf{X}}$ is a "linearized" image, that is, an image which is one dimensional, but where the two dimensional relations between elements are known, then the columns of $\underline{\mathbf{B}}$ will also contain the same relations, again transformed.

A "linearized" image is actually a familiar concept, as a computer stores an image linearly, but uses a predetermined addressing scheme to access the image in a two dimensional fashion. Two examples of this are row major order, where each row of the image is stored sequentially, or column major order, where each column is stored sequentially. Regardless of the method used, each row of the matrix $\underline{\mathbf{X}}$ when created using linearized images, represents the temporal evolution of an image pixel.

To therefore analyze image sequences the images are first linearized, then the KLT applied. The basis images are extracted from the basis vectors by "undoing" the linearizing transformation.

Note also, that since the KLT is a linear transform any linear operation performed upon the data set before analysis may be performed afterwards on the basis images and coefficients. This is particularly useful for techniques like interpolation. For example, interpolating from the 31 measurement electrodes to 64x64 images can be performed only on the basis images, which in this case will amount to no more than 31, rather than interpolating all images analyzed, which may amount to several hundred. Other transformations such as Laplacian filtering for detection of current generators, and power spectrum estimation may also be performed after KLT analysis with a similar savings in computation.

In Chapter 2 it was stated that by appropriate normalization the interpretation of $\mathbf{\underline{B}}$ and $\mathbf{\underline{K}}$ could be reversed. If such a process were performed each row of $\mathbf{\underline{K}}$ could be thought of as a basis function, where the first basis function represents the best least squares fit to the temporal evolution of all pixels. Each basis function thus corresponds to a time series signal (of length T) while each column of $\mathbf{\underline{B}}$ contains the appropriate weighting factor for each basis signal. In the case of analyzing a set of electrode recordings this is generally the interpretation given (Glaser and Ruchkin, 1976), while for images the former interpretation is used.

3.3 Program Notes

The programs to analyze the spatial EEG are written in VAX FORTRAN, on the Department of Applied Sciences in Medicine VAX/11-750 computer. Most mathematical computations are performed using EISPACK routines or are calculated with an FPS AP/120B array processor, using a 28 bit mantissa and a 10 bit exponent, attached to the VAX.

Eigenanalysis using both EISPACK or the FPS array processor consists first of tridiagonalizing the data matrix using the Householder method. The eigenvalues and vectors of the data matrix are then found by applying the QL algorithm to the products of tridiagonalization. A description of these algorithms can be found in Smith et al. (1976); a detailed outline is not provided here as it would be beyond the scope of this thesis.

While the programs written for analyzing the spatial EEG sequences are straightforward, and will not be discussed in detail, several key points should be mentioned. After data collection the 31 channel EEG data (raw images) are converted into 64x64 spatial EEG maps by linearly interpolating the 4 corner electrode values, completing a 5x7 grid, and then to 64x64 pixels by either linear or bicubic spline interpolation. As was stated earlier, these techniques may be applied after KLT analysis to the basis functions rather than to the entire image set in order to save computation time.

The programs allow for selection of raw (31 channel data), or image (64x64) sequence data as input. After KLT analysis when raw images are used as input, the output basis images may be raw as well, or image size (64x64), created using either linear or bicubic spline interpolation, and for the latter, further processed by gradient or Laplacian filtering. Before analysis the user may prenormalize the input data using one of 5 modes: each image may be made zero mean; each pixel sequence (electrode sequence when raw images are used) may be made zero mean; each image may be normalized to have unit power (i.e. sum of squares is 1.0) after removal of the temporal mean of each pixel; the autocorrelation function of the image sequence may be multiplied by a prenormalizing pattern classification matrix (discussed in Chapter 4); and finally no normalization may be performed.

As was stated in Chapter 2 the basis functions are computed from either autocorrelation matrix, $\mathbf{R}$ (NxN) or $\mathbf{Z}$ (TxT). The number of non-zero basis functions is the smaller of the number of images used (T) or the image size (N) used. Since T rarely exceeds 1200 (10 seconds of data) it is more economical to perform eigen analysis on $\mathbf{R}$ when the image size is small (i.e. when raw data is used), or on $\mathbf{Z}$

when the number of images to analyze is small (i.e. when interpolated images are used). To save computation time the program automatically selects whether to perform eigen analysis on $\underline{\mathbf{R}}$ or $\underline{\mathbf{Z}}$, depending upon the size of N and T.

Several other features included in the analysis programs include the colour display of the basis images, and the option to save the basis images and the associated coefficient sequences for further investigation. The ability to reconstruct images using a varying number of basis images and coefficients and visually compare them, as well as compute their correlation, is also available.

The data for the wireframe and 2D grey scale images displayed in this thesis were transferred to a Macintosh™ computer, where the images were generated. The programs to create these images were written by the author in the C programming language, using a Macintosh™ computer.

3.4 Pattern Representation of the EEG Using the Karhunen-Loeve Transform Within Subject Records

The first, and simplest, application of the KLT to image sequence analysis is to determine if, within an individual, certain topologies of electrical activity are characteristic of the spatial EEG. That is, can certain images be added together linearly to represent the spatial EEG? Since there are 31 electrodes (independent measurements) the maximum number of basis images required when analyzing the "raw" maps, is at most 31. If the spatial EEG is a totally random process then all 31 basis images are required, each accounting for about the same power, but if the process has spatial correlation then fewer basis functions are required to represent the data set.

Recall that the measure of the total power which a certain basis function represents is given by its corresponding eigenvalue, or by the mean squared value of its coefficient sequence (row of matrix **K**). The percentage of power that each basis function represents is thus its corresponding eigenvalue divided by the sum of all eigenvalues. If instead of using the eigenvalues this relationship is expressed in terms of the coefficients and data elements (via equation 2.3.1.8) the following equation emerges (after cancellation of the 1/T terms)

$$\begin{matrix}\text{\% Average Power Contributed} \\ \text{From the } i^{\text{th}} \text{ Basis Image}\end{matrix} = \frac{\sum_{t=1}^{T} k_{it}^2}{\sum_{i=1}^{N}\sum_{j=1}^{T} x_{ij}^2} * 100\% \qquad (3.4.1)$$

This is a useful value in comparing the effectiveness of basis images between various subjects, and even within the same subject, since it removes the effects of the signal power from the computations. Note also that in the form written above the eigenvalues are not required to compute the ratio, not all basis function coefficients are required, and the number of images used to compute these percentages can be variable. In effect then, Equation 3.4.1 measures the normalized projection of data onto a basis function. It is from this latter point that the usefulness of the above measure resides: the effectiveness of basis images in representing data not used in computing those basis images can be measured after computing only the appropriate coefficients.

Using this property of Equation 3.4.1 leads naturally to asking if basis images computed from one section of EEG can be used represent other data segments, possibly even from different cognitive states? Since the dimensionality of the signal space does not change through time the basis functions found from one segment may be used in another segment, though they may not fit optimally in the least squares sense. That is, the data may be exactly represented by using the original basis images, but the the basis image accounting for the largest percentage power from the original segment may not account for the largest percentage in other segment. Moreover, there is the problem of how to compute the coefficients in order to determine the percentage of average power the basis function represents.

The first of these two problems arises from the fact basis functions computed from a data set are optimal only for that data set. Recall that optimal means that the first basis image and coefficient sequence account for the largest percentage of power of the original data set with the rest of the basis images each accounting for a successively smaller percentage. When fit to new data, this will most likely not be the case; in other words, the order may not strictly be descending. This may not be so bad, though, if in general the descending pattern is maintained and if the first several basis images still account for the largest portion of the power. If, on the other hand, the distribution becomes quite random or the percentage weights shift considerably to other basis images there will be little spatial cohesiveness in the spatial EEG. The point at which such a shift occurs, if at all, may be used as a useful measure of the spatial "stationarity" of the EEG.

To be able to answer such questions demands that the coefficient sequence corresponding to each basis function be computed. For the original data set when the basis images are known, the coefficient matrix can be computed using Equation 2.3.3.25. This operation is simply a projection of the data set onto the basis image space, where, for the original images, it is known that the basis space will optimally span the data. The question thus arises, can the same matrix operation be used when the basis space is not known to optimally span the data (but known to be of the same dimensionality)? To answer this, a least squares fit between the basis images and the new data set is used.

To simplify the mathematics, and without loss of generality, assume that the data images and basis images are "linearized", as was discussed earlier. It is desired to minimize the mean squared error between the i^{th} basis image and the j^{th} data image by adjusting the basis images coefficient. The value may be found by minimizing

$$\begin{array}{c}\text{Squared error in using } i^{th} \text{ basis image} \\ \text{fitting to } j^{th} \text{ data image}\end{array} = (\underline{\mathbf{E}}_j - k_{i,j}\,\underline{\mathbf{b}}_i)^2 \qquad (3.4.2)$$

where $\underline{\mathbf{E}}_j$ is the j^{th} data image, $k_{i,j}$ is the coefficient of $\underline{\mathbf{b}}_i$, the i^{th} basis image.

Equation 3.4.2 is only valid if the basis images form an orthogonal set, otherwise the $\underline{\mathbf{E}}_j$ would have to represent the residual error "image" between the original image and the previous i-1 basis images (for i=1 $\underline{\mathbf{E}}_j$ would be the original image). $\underline{\mathbf{E}}_j$ may represent the data image because subtracting only orthogonal components from the original image does not change the projection onto different orthonormal basis images. The following derivation will refer to $\underline{\mathbf{E}}_j$ as the original image since, in this case, it is known that the basis images are orthonormal.

Performing the partial differentiation of equation 3.4.2 yields

$$\frac{\delta}{\delta k_{i,j}} \left(\underline{\mathbf{E}}_j^t \underline{\mathbf{E}}_j - k_{i,j} \underline{\mathbf{E}}_j^t \underline{\mathbf{b}}_i - k_{i,j} \underline{\mathbf{b}}_i^t \underline{\mathbf{E}}_j + k_{i,j}^2 \underline{\mathbf{b}}_i^t \underline{\mathbf{b}}_i \right) = \\ - \underline{\mathbf{b}}_i^t \underline{\mathbf{E}}_j - (\underline{\mathbf{b}}_i^t \underline{\mathbf{E}}_j)^t + 2\, k_{i,j} \underline{\mathbf{b}}_i^t \underline{\mathbf{b}}_i \qquad (3.4.3)$$

Noting that the last bracketed term is a scalar (and thus the transpose operation has no effect) the first two terms may be grouped together. Setting this result to zero and solving for $k_{i,j}$ gives

$$k_{i,j} = \frac{\underline{\mathbf{b}}_i^t \underline{\mathbf{E}}_j}{\underline{\mathbf{b}}_i^t \underline{\mathbf{b}}_i} \qquad (3.4.4)$$

Since the basis images form an orthonormal set the denominator is unity and thus the coefficients are given by

$$k_{i,j} = \underline{\mathbf{b}}_i^t \underline{\mathbf{E}}_j \qquad (3.4.5)$$

Therefore, the coefficients are computed by projecting the new images onto the basis functions, with the assurance that the mean squared error in using each basis image will be minimum. This does not guarantee, though, that the mean squared error for the first basis image is the minimum possible, or even minimum over all basis images. For the data set from which the basis images were extracted the mean squared error for each basis image is a minimum, but this is not necessarily so for any other data set.

Once the basis images and coefficients have been computed it is desirable to have a different measure, other than mean squared error, which relates how similar the reconstructed images and original images are. The correlation between the two images is such a yardstick since it measures similarity in image shape, independent of their magnitudes. The following gives the normalized correlation between two linearized images, i.e. vectors,

$$\text{correlation} = \frac{\sum_{i=1}^{N}(o_i - \bar{o})(c_i - \bar{c})}{\sqrt{\sum_{i=1}^{N}(o_i - \bar{o})^2 \sum_{i=1}^{N}(c_i - \bar{c})^2}}$$

For computational efficiency the following equivalent form is used here for computing the correlation,

$$\text{correlation} = \frac{\sum_{i=1}^{N} o_i c_i - \frac{\sum_{i=1}^{N} o_i \sum_{j=1}^{N} c_j}{N}}{\sqrt{\left(\sum_{i=1}^{N} o_i^2 - \frac{\left(\sum_{i=1}^{N} o_i\right)^2}{N}\right)\left(\sum_{i=1}^{N} c_i^2 - \frac{\left(\sum_{i=1}^{N} c_i\right)^2}{N}\right)}} \qquad (3.4.6)$$

where c_i is the ith element of the reconstructed image, o_i is the ith element of the original image, $\bar{o}$ and $\bar{c}$ are the mean values of each image, and N is the number of elements in each image.

Values of correlation range from -1 to 1, the larger the absolute value the more similar the images. A correlation coefficient of 1 indicates that the images are the same, a value of -1 means that the images are the same within a negative scalar. A correlation coefficient of zero indicates that the images are dissimilar. It must be noted that two orthogonal images, that is, their pixel by pixel sum is zero, are not necessarily uncorrelated. In order to be uncorrelated one or both of the images must also be zero mean, a fact clearly shown in Equation 3.4.6.

Once a correlation has been computed it is reasonable to ask how closely the value represents the true correlation, ρ, between each of the processes. If only a single correlation is computed, that is, between an original 31 point measurement image and its reconstructed version, Fischer's (1921) approximation is used to determine the confidence bounds. When correlations are averaged over several images, a value known here as the average reconstruction correlation (ARC), the Student t distribution is used (Cooper, 1969).

Based upon averaging 1200 images, the number used throughout this thesis, and a 99% confidence interval the variation around each measured average correlation value is +/-0.074 σ, where σ is the standard deviation of the 1200 correlation values. It is easy to see that even for a large σ the ARC is very accurate. Using a σ of 0.1 (larger than any measured on the data in this thesis) the average correlation values are significant within approximately 0.01. Therefore, any change of the ARC of more than 0.01 is significant using 99% confidence bounds. Unless otherwise stated, this value applies to all ARC values listed throughout this thesis.

As well as correlation, power values are averaged over 1200 images, and thus the 99% confidence interval is also +/-0.074 σ. Again a ρ of 0.1, larger than any measured on the data used herein, gives a confidence interval of +/-0.01, or +/-1%. This value applies to all power values computed using 1200 images listed throughout this thesis.

Another question which arises is how significant are the correlation values? Gibra (1973) provides a method for estimating their significance, again using the Student t distribution. Based on 31 measurement points (electrodes), that is 29 degrees of freedom, and a 99% confidence interval, any correlation value over 0.456 is significant. In other words, the null hypothesis ($\rho = 0$) can be rejected with 99% certainty.

At this point it is possible to investigate temporal sequences of spatial EEG maps by applying the KLT to generate basis images and coefficients. It is also possible to see how well the basis images computed from one segment of images "fits" another by applying Equation 3.4.5 to compute the coefficients, and Equation 3.4.1 to measure the "effectiveness" of each basis image.

In the following section the results from the analysis of one individual will be given in detail, followed by the results from applying the technique to several subjects. The data used in the following investigation has been acquired as described in Section 3.1.

3.4.1 Detailed Study - Methods & Results

In this section the KLT is applied to an individual's spatial EEG to determine how well the computed basis images represent the original spatial maps. To fully answer this, three sub-questions are asked, namely: how many basis images are required to adequately reconstruct the original images (from which the basis images were computed), how many data images are needed to compute these basis images in order to get a reasonable estimate of the basis images, and can basis images from one section of data be used to represent data images from other segments of the recording session? In this section, these questions are answered and discussed in detail using the data from one subject.

The following results were taken from two minute artifact free records of EEG obtained from an individual whose alpha activity in the EC state was very strong, and whose EP data had strong activity as well. The analysis concentrated on using EC data, although EP data was also used. Data from this subject is denoted by Subject 1 EC, or Subject 1 EP, depending upon the record used.

Table 3.4.1.1 shows the percentage of average power (via Equation 3.4.1) represented by each basis image after applying the KLT to the first 1200 (=10 seconds of) raw images (i.e. 31 point electrode values only) of the eyes closed record. Before this analysis the temporal average per electrode over the first 10 second period was removed from each electrode value to avoid unwanted effects such as amplifier drift and noise.

Basis Image	%Power Represented by Basis Image	Cumulative Power
1	52.44	52.44
2	27.79	80.23
3	8.12	88.35
4	4.52	92.87
5	2.22	95.09
6	1.57	96.66
7	0.75	97.41
8	0.65	98.06
9	0.36	98.42
10	0.30	98.72

Table 3.4.1.1
Distribution of Basis Function Power For The First 10 Basis Images Computed From A 10 Second of Image Sequence - Subject 1 EC

Since the mean value was removed from each electrode it is possible to create a "mean image", that is, one in which each pixel represents the value subtracted from the data image (or electrode, in this case) before KLT analysis. Figure 3.4.1.1 displays a 64x64 bicubic spline interpolated representation of the mean electrode values, giving a "mean image". It is important to note that because subtraction of these mean values was done prior to analysis, when reconstructing the images the mean image must be summed together with the product of the basis images and coefficients.

The first three basis functions are shown in Figures 3.4.1.2* through 3.4.1.4*, respectively. The images shown are 64x64 pixels, interpolated using the bicubic spline from the 31 measurement values, with the labeling showing the orientation of the head. Note that the basis images are dimensionless.

---

*The orientation of all images shown in this thesis is such that the subject is facing the right hand side of the page with the viewer above the subject, unless otherwise indicated by the labeling.

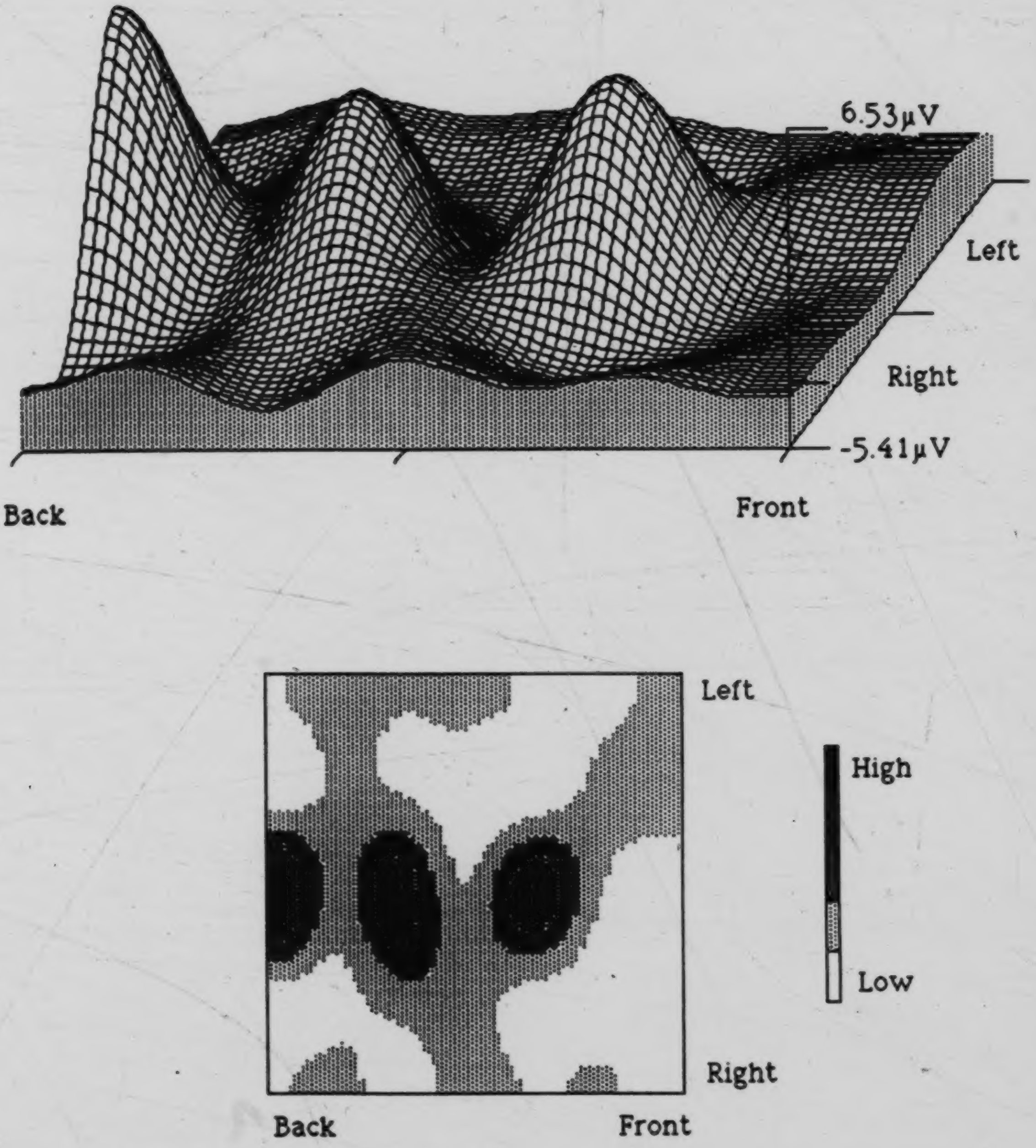


Figure 3.4.1.1
Mean Image Computed From First 10 Seconds of Image Sequence - Subject 1 EC

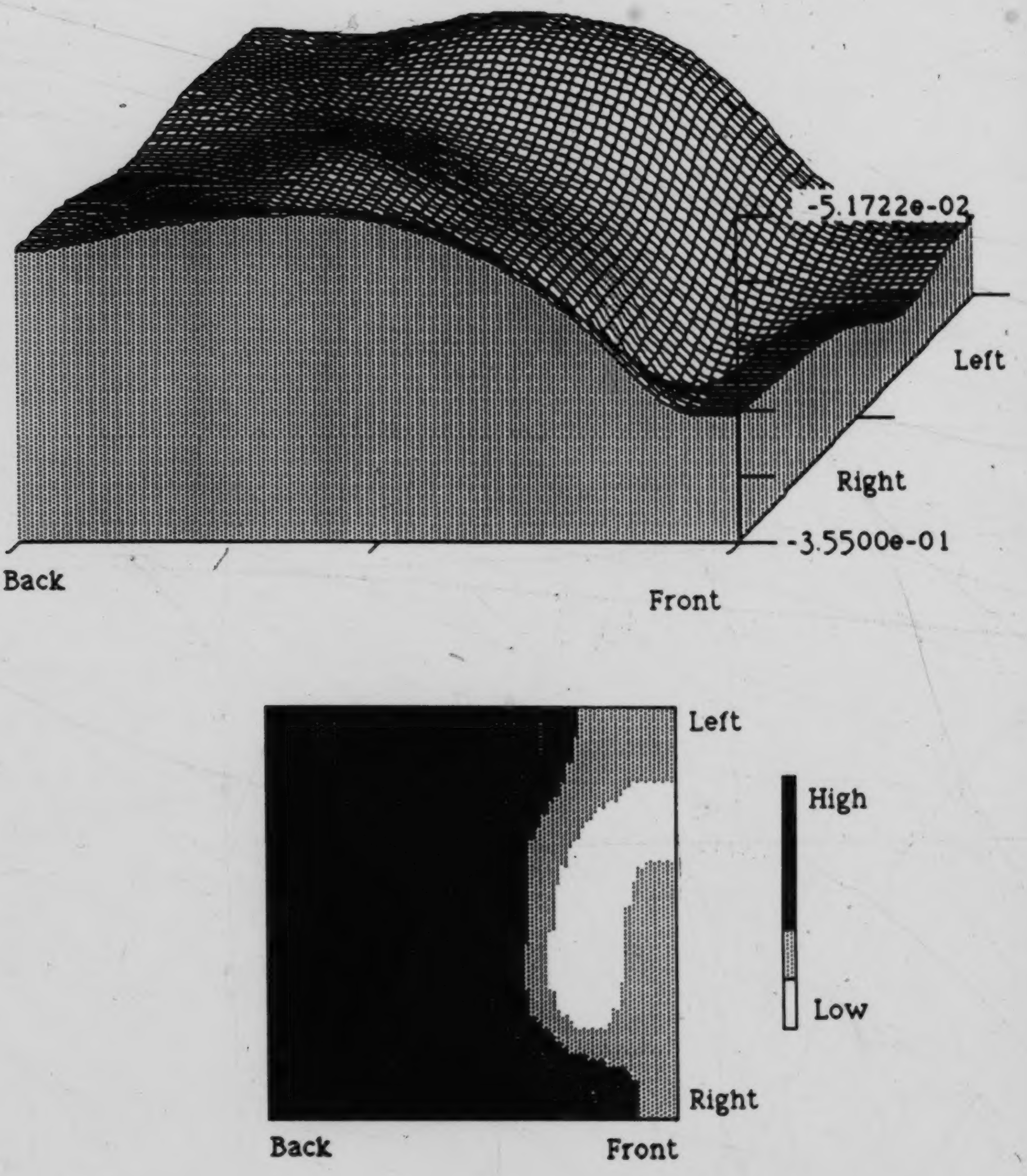


Figure 3.4.1.2
Basis Image 1 Computed From First 10 Seconds of Image Sequence - Subject 1 EC

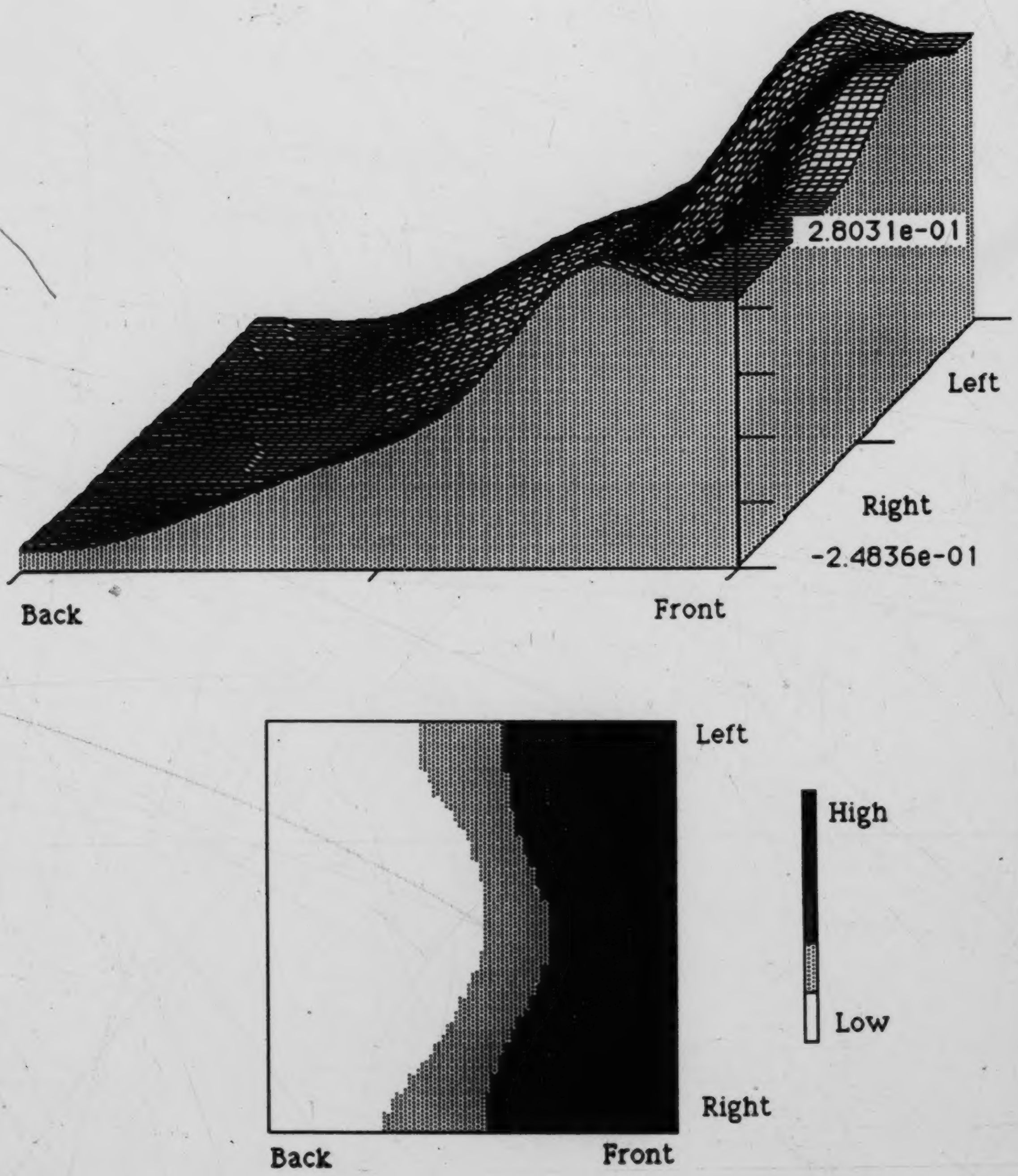


Figure 3.4.1.3
Basis Image 2 Computed From First 10 Seconds of Image Sequence - Subject 1 EC

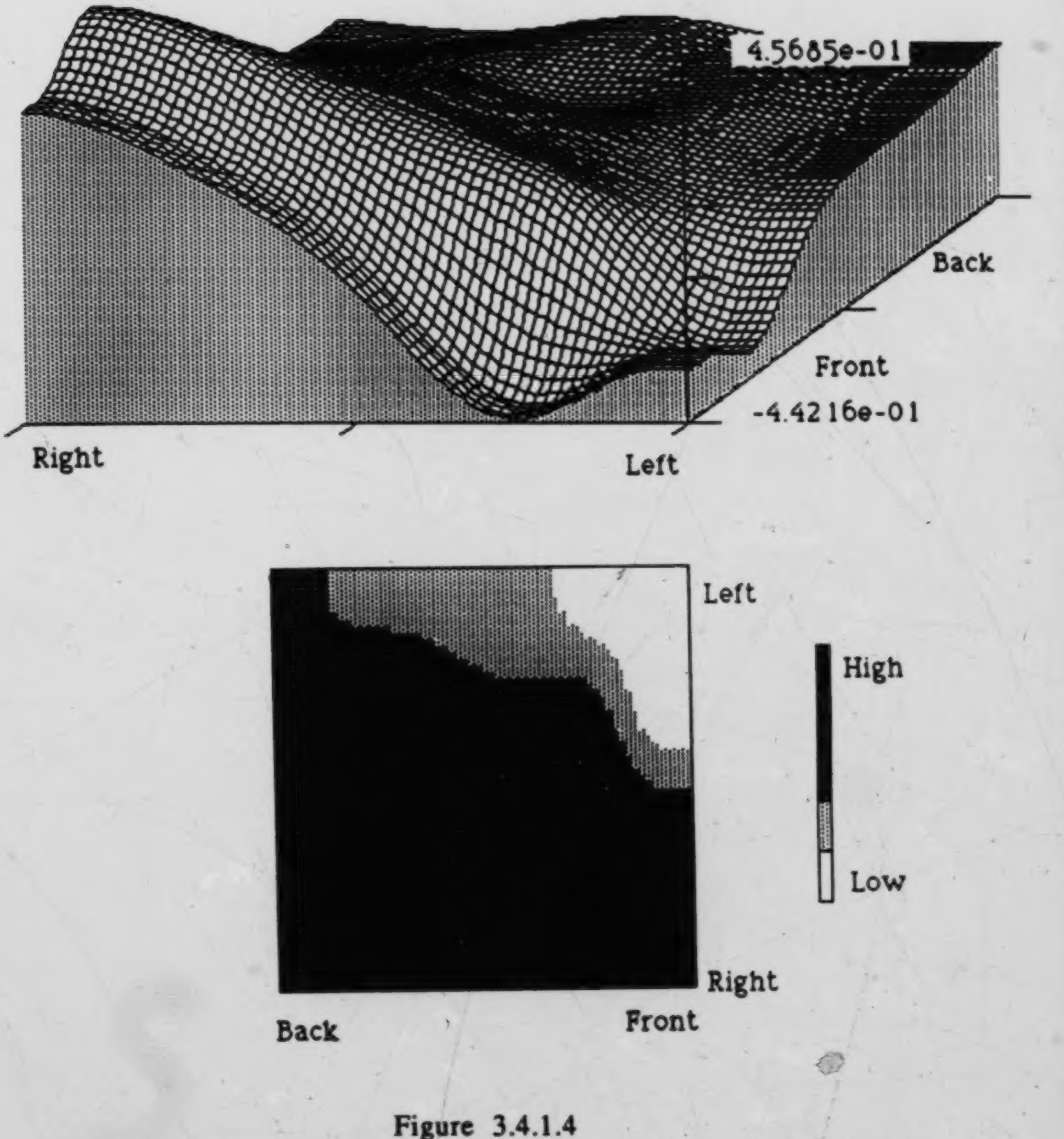


Figure 3.4.1.4
Basis Image 3 Computed From First 10 Seconds of Image Sequence - Subject 1 EC

The coefficient sequence of the first two (out of 10) seconds, for the first basis image is shown in Figure 3.4.1.5. Since each channel recording was made zero mean for the analysis period the coefficient sequence is also zero mean over the same period (even though for the first two seconds shown the sequence is not zero mean).

Using the first five basis images, out of the possible 31, the correlation* between the original and reconstructed images was measured and averaged over the same 1200 images used to compute the basis images. This measure, called the average reconstruction correlation (ARC), was 0.93. Figure 3.4.1.6 shows the actual correlation values for first 2 seconds of the 10 second segment, as well as the amount of power the 5 reconstructed images represented. Recall that ARC values are significant to within +/-0.01, and that individual correlation values are significant if over 0.456, both using 99% confidence bounds.

---

* All reconstruction correlation values used in this thesis are based on correlating between the 31 point measurement values, not on the 64x64 interpolated representations of the images.

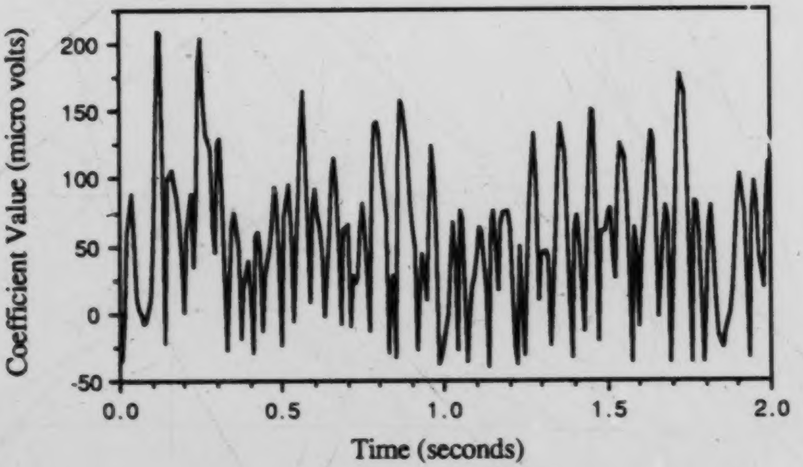
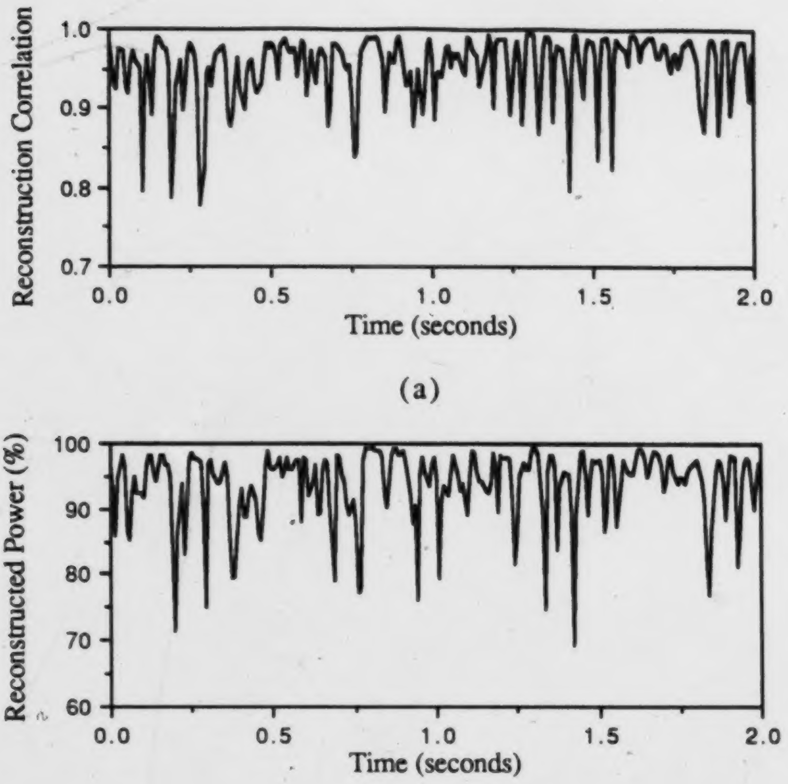


Figure 3.4.1.5

Coefficient Sequence of Basis Image 1 during first 2 Seconds of Analysis - Subject 1 EC



(a)

(b)

Figure 3.4.1.6
Reconstruction Parameters Using First 5 Basis Images For First 2 Seconds of Subject 1 EC Data Record (a) Correlation (b) Power

Figure 3.4.1.7 shows a typical data image from the first 10 seconds of the EC record, and a reconstructed version of that image based on the first five basis images. As before, each image is a 64x64 bicubic spline interpolated version of the 31 measurement set. The correlation between these two images is 0.98, with 99% confidence limits of 0.97 and 0.99. The power represented by the reconstructed image is 96.6%, with 99% confidence limits of 95.6% and 97.7%. For comparison purposes both of the wireframe representations are shown in the same orientation.

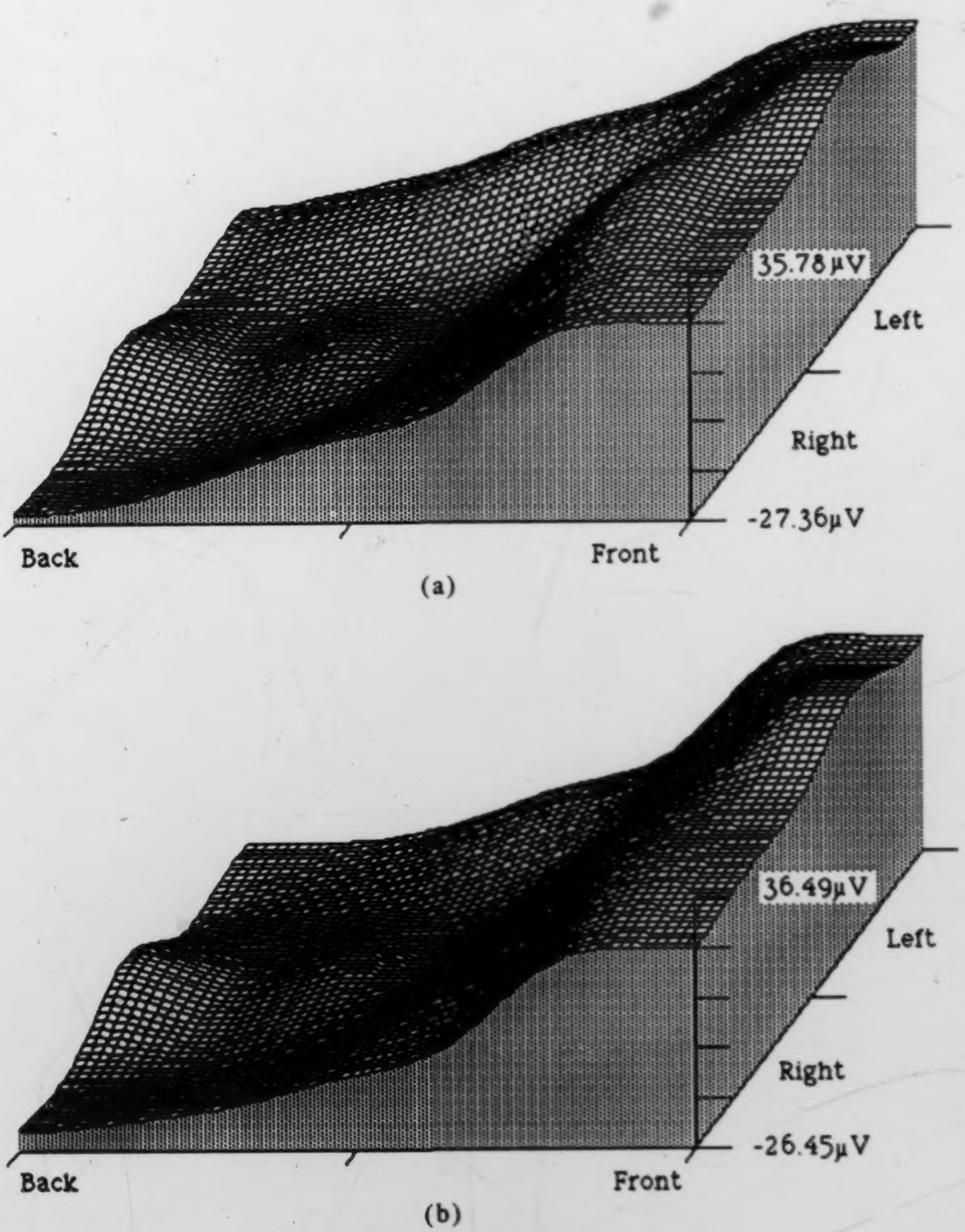


Figure 3.4.1.7

Image Reconstruction Using the First Five Basis Images : (a) Original Image (b) Reconstructed Image

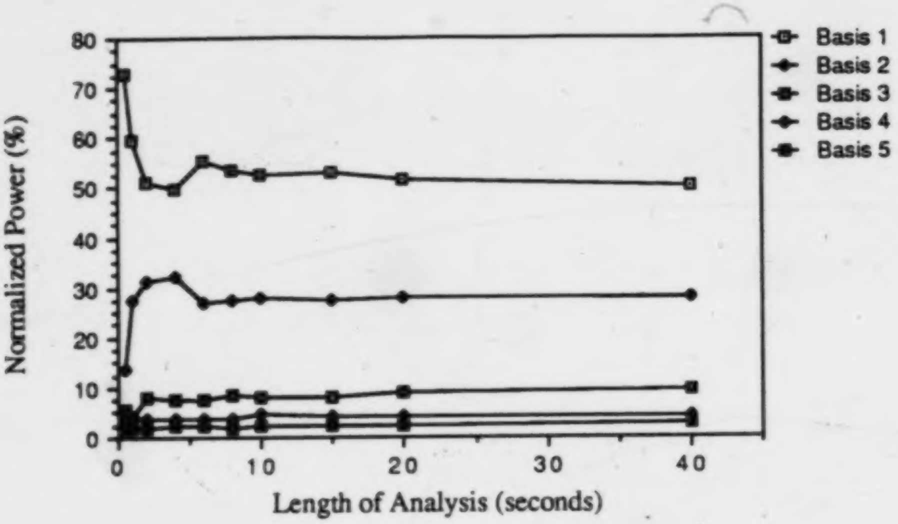
The choice of using 10 seconds of data to compute the basis images, at present, seems quite arbitrary. Why should such a value be used, why not 120 images to save computation time, or 12000 images to ensure that the second order statistics of the EEG have been fully captured in the autocorrelation matrix?

In order to answer this Figure 3.4.1.8 displays the normalized power represented by each basis image when the amount of data was varied (all data sets started from the first image in the data record). Figure (a) shows the results when the optimal basis images were used, that is, ones computed from the actual data, while Figure (b) shows the results when basis images computed from the first 10 seconds of the record were used.

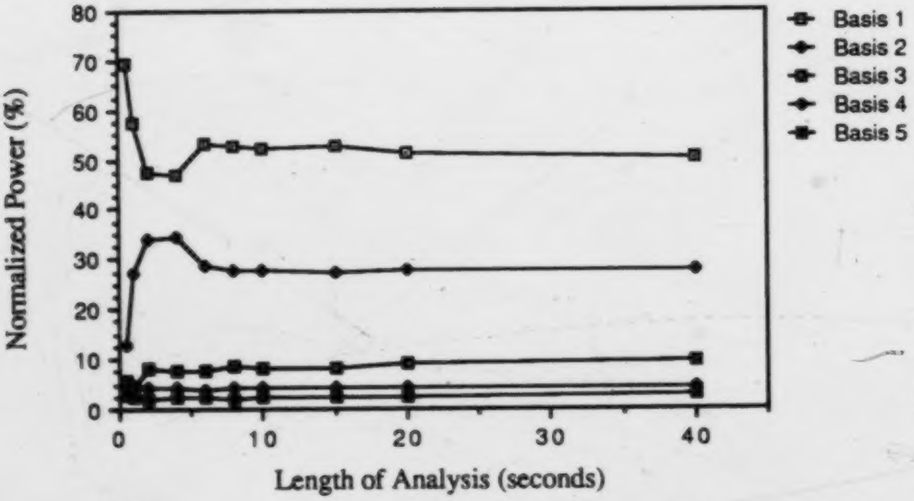
Using the basis images computed from the first 10 seconds of the eyes closed record it is also possible to compute coefficients, via Equation 3.4.5, that fit them to other segments of the record. This can also be viewed as projecting the data onto the space defined by the basis images. With the coefficients the effectiveness of each basis image can again be measured with Equation 3.4.1.

Figure 3.4.1.9b gives the power represented by the first five basis images from the first 10 second section of data, for other 10 second data sets, starting at various points throughout the record. For comparison, the power represented by the optimal basis set for each such period was also computed and is shown in Figure 3.4.1.9a. The numbers above each group of bars give the total power represented by the 5 basis images. Table 3.4.1.2 provides the ARC for both of these analyses.

As always, each electrode sequence was made zero mean for the period of analysis.



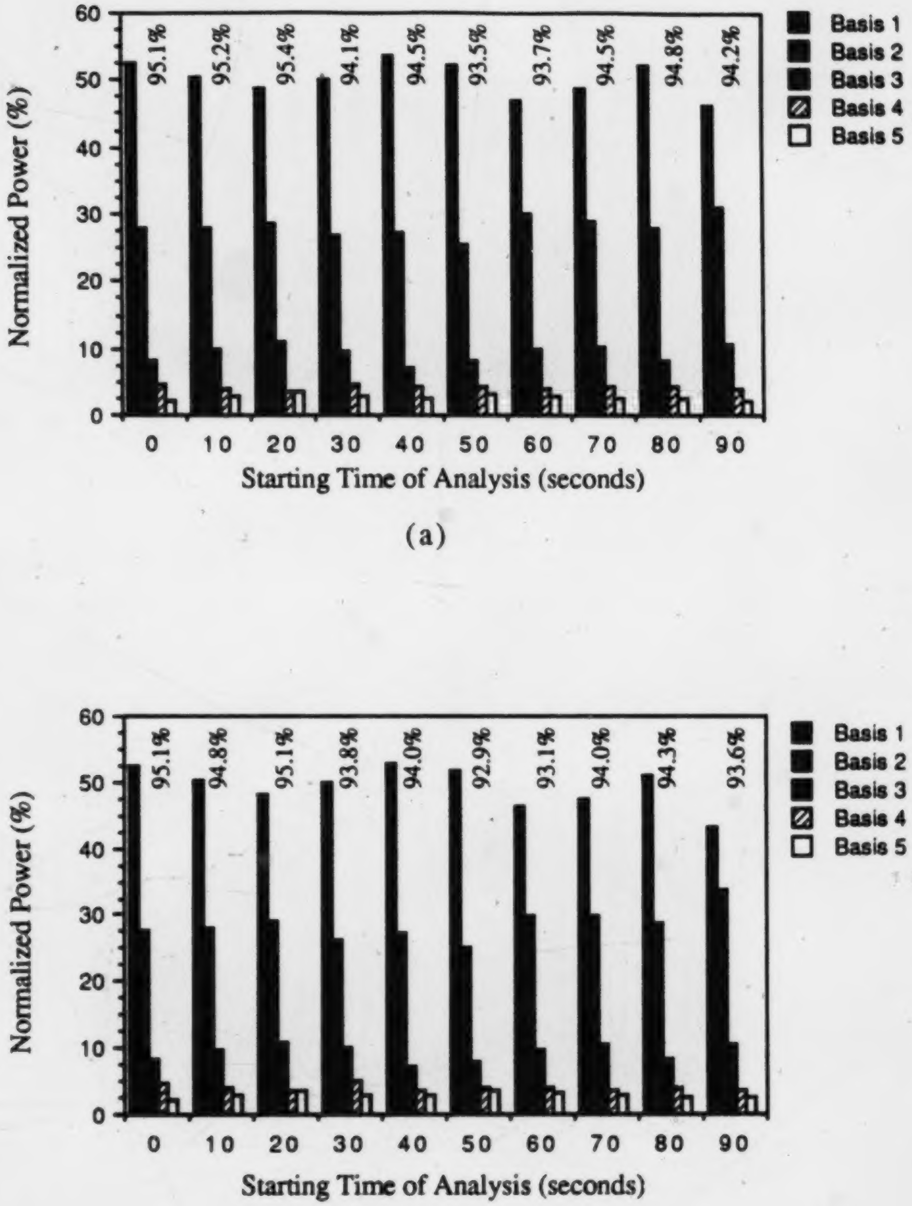
(a)



(b)

Figure 3.4.1.8

Basis Image Power Representation For Varying Lengths of Data - Subject 1 EC : (a) Using Optimal Basis Images (b) Using Basis Images Computed from First 10 Seconds of Record



(a)

(b)

Figure 3.4.1.9

Distribution of Basis Image Power For 10 Second Segments of Data Record - Subject 1 EC : (a) Using Optimal Basis Images (b) Using Basis Images Computed From First 10 Seconds of Record

Starting Time of Analysis (seconds after T=0)	Average Reconstruction Correlation	
	Optimal	Using Basis Images From First 10 Seconds
0	0.93	0.93
10	0.93	0.93
20	0.93	0.93
30	0.92	0.92
40	0.92	0.92
50	0.91	0.90
60	0.91	0.90
70	0.92	0.92
80	0.92	0.92
90	0.92	0.91

Table 3.4.1.2
Average Reconstruction Correlation For 10 Second Segments of Data Record - Subject 1 EC

Eyes open (EP) data, which does not typically contain as strong an alpha rhythm component as EC data, was also analyzed. Rather than repeat the entire analyses performed above, only a subset was done. The power represented by the first 5 optimal basis images for varying lengths of data is shown in Figure 3.4.1.10. The ability of a set of basis images computed from one section of data (here again, the first 10 second segment) to represent images from other sections of the data record is shown in Figure 3.4.1.11, with the ARC values obtained from that analysis listed in Table 3.4.1.3.

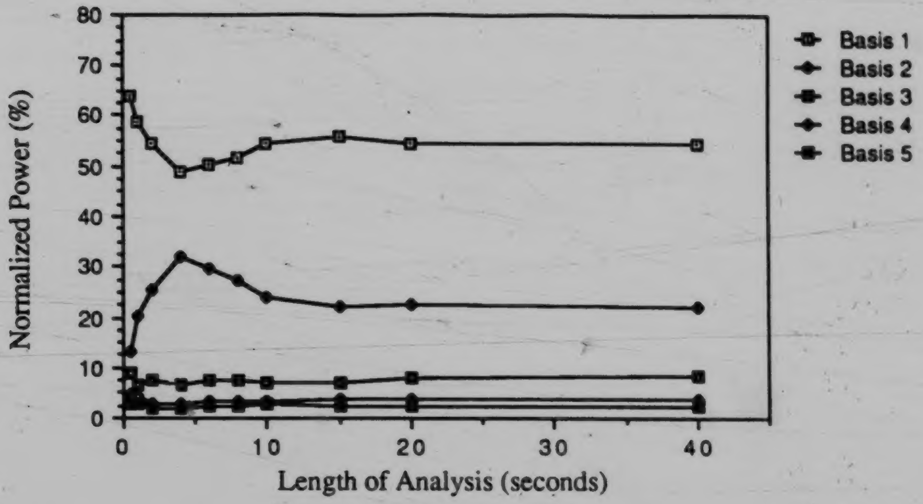


Figure 3.4.1.10
Power Represented by Optimal Basis Images for Varying Lengths of Data - Subject 1 EP

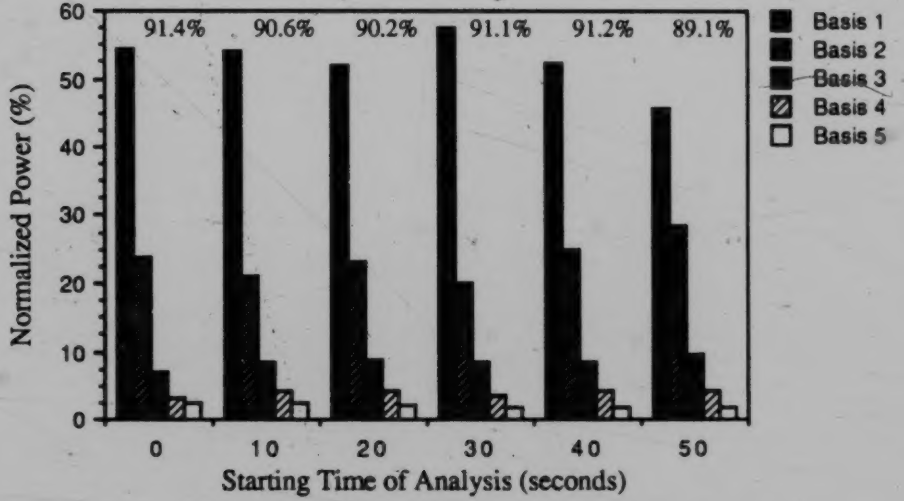


Figure 3.4.1.11
Distribution of Basis Function Power when Basis Images Computed From the First 10 Seconds are Applied To Other 10 Second Segments of Data Record - Subject 1 EP

Starting Time of Analysis (seconds after T=0)	Average Reconstruction Correlation
0	0.89
10	0.88
20	0.88
30	0.89
40	0.89
50	0.88

Table 3.4.1.3
Average Reconstruction Correlation Corresponding to Figure 3.4.1.11 - Subject 1 EP

3.4.2 Detailed Study - Discussion

Table 3.4.1.1 shows that the first five basis images account for over 90% of the average power of the original data set. Since the signals have been made zero mean this figure also represents the amount of variance represented by the basis functions. These results, along with the high reconstruction correlation measured, indicate that the scalp electrical activity of this subject is represented by distinct spatial patterns which combine linearly to form the observed data. Whether or not these basis images are physiologically meaningful is not investigated here, as it is beyond the scope of this thesis.

Another point is that with the KLT data compression may be performed on the EEG. For example, if the first five basis images and associated coefficients are stored instead of the original data, only about 17% of the original storage memory is required. For sequences of images using megabytes of storage this amounts to considerable savings, with little associated error both in signal power, measured using the mean squared error, and shape, measured using correlation.

The mean image, shown in Figure 3.4.1.1, indicates that the d.c.

level of three electrodes in the back-center of the scalp is higher than the other electrodes. Since these average values should be zero, because of the a.c. coupling of the electrodes to the recording equipment, they are unrelated to the actual EEG and are thus not meaningful. By subtracting the mean values more valuable shape information is therefore captured by the KLT. It is for this reason that all analysis performed in this thesis removes the temporal averages from each electrode value before any further processing takes place.

The basis images shown in Figures 3.4.1.2 through 3.4.1.4 exhibit interesting properties. Basis images 1 and 2 indicate that much of the scalp electrical activity is represented by a simple front/back pattern. The third basis image contains side to side activity, but only represents about 8% of the power in the original images.

It is important not to discount the value of the higher number basis functions, even though they do not represent much of the power in the original images. This is because they represent the fine details of shape, rather than the gross features of the data. In fact, in some pattern classification methods the lower numbered basis functions are used as discriminatory features, since in differentiating between various pattern classes detail, rather than the coarse structure is more important (Tou and Heydorn, 1967).

Returning to the data at hand, a portion of the coefficient sequence for the first basis image is shown in Figure 3.4.1.5, and exhibits significant periodicity. As well as showing periodicity it is worth noting that the coefficient sequence is zero mean over the analysis period, since the data itself was made zero mean. It is important to recall that the coefficient sequences corresponding to the basis images are orthogonal as are the basis images themselves, and thus through appropriate normalization can be considered basis functions. In such a case the basis functions do not represent spatial basis images, but rather basis functions which account for the temporal activity at each measurement electrode. Hence the periodic activity seen in the coefficient sequence relates to the temporal characteristics (predominantly alpha band activity) seen at the electrodes.

There is a disadvantage in viewing the KLT products in such a way since these temporal basis functions (while numbering no more than 31 in this case) have varying lengths depending upon the amount of data analyzed, and are thus difficult to compare. One way to avoid this problem is to use the Fourier Transform of the coefficient sequence to characterize temporal basis functions, although this avenue is not pursued here.

Figure 3.4.1.6a shows the correlation between the original images and reconstructed versions of those images, created using the first five basis images, for the first two seconds of the 10 second analysis period. Figure 3.4.1.6b is similar, except that power instead of correlation is measured. Generally the correlation and power are high, although there are times when they are both significantly lower than average (0.93 for correlation, 95.1% for power). These moments may indicate when the spatial EEG is going through transitory phases, or when it cannot be represented linearly. This may coincide, for example, with the rise and fall of alpha activity. In general, though, the figure shows that using the first five basis images provides reasonable shape and power reconstruction for all data images.

To demonstrate how well images are reconstructed Figure 3.4.1.7 shows an original and reconstructed image. The reconstructed image shown is created from the first five basis images and the mean image from the analysis period, and correlates 0.98 with the original image. Before comparing the two images it is interesting to note the three "bumps" located along the mid-line, in approximately the same locations as those shown in the mean image in Figure 3.4.1.1. This again reinforces the idea that removing the mean image before KLT analysis is useful for the basis images to capture shape information independent of electrode d.c. bias. This also shows that the mean values are relatively small in relation to the signal itself, and thus any drift in an amplifier's d.c. bias is not significant.

In general, the shape and details of the original image are well preserved in the reconstructed image. These include the front/back pattern and even the "bumps" associated with the d.c. electrode bias. Some differences do exist. For example, the reconstructed image has a higher maximum (by approx. 2%) and higher minimum (by approx. 3.4%) value than that of the original image. Overall, the reconstruction is surprisingly good considering that only 5 out of a possible 31 basis images are used in the reconstruction.

Figure 3.4.1.8 shows the normalized average image power represented by the first five basis images for varying lengths of analyses, for both the optimal basis images (computed from the data itself), and for those computed from the first 10 seconds of the record. Both results are very similar, the only significant difference being that in the optimal case at short data lengths the power in basis image 1 is higher, and that of basis image 2 lower, than the values using the precomputed basis images. For longer analysis lengths, starting at around 8 seconds, both results are extremely close. The similarity between these two results demonstrate that basis images computed using the first 10 seconds of the record are representative of varying lengths of data as short as 0.5 seconds, and as long as 40 seconds. This also indicates that the basis images themselves do not significantly change within these limits.

It is interesting to note that when 8 seconds or more of data are used in the analysis the power represented by the basis images remains relatively constant. Therefore the second order statistics of the EEG, captured in the autocorrelation matrix, are not changing significantly for these lengths, while for shorter lengths the figure illustrates large variations in power representation (for basis image 1 and 2 as much as 20%), and that for very short intervals the power is concentrated in basis image 1. This latter point indicates that over these short intervals less spatial variation is present than over longer intervals, i.e. fewer spatial patterns exist in short data segments.

Using the results discussed above the reason for using basis images computed from 10 seconds of data should now be clear: they are representative of all lengths of data, involve fewer computations than using the entire record, and are not subject to fluctuations in the statistics of EEG as those from shorter lengths. For these reasons most data analyzed throughout this thesis is based on using 10 second segments of data.

Figure 3.4.1.9 shows the power represented by the first five basis functions computed from various 10 second segments of the EC record. The figure also shows the power represented by the first five basis images computed from the first 10 second record when the same 10 second segments as above, are projected on to them. For the latter, it is worth noting that the order of importance of each basis image is preserved, that is, the first basis image always accounts for the largest amount of power (minimum least squared error), and likewise for the other four basis images. Moreover, the results using the non-optimal basis images are very similar to the ones for the optimal images (computed from each data segment), even for data separated by over 60 seconds. Table 3.4.1.2 confirms this assertion by showing that the average reconstruction correlation for each projection is consistently high, and near optimum, regardless of the segment of the record used. These results demonstrate, again, that the basis images computed from the first 10 second record, are representative of the entire record in general, in both shape and importance.

The two main ideas investigated above, that of analyzing varying lengths of data to determine the characteristic basis images of the spatial EEG, and that of seeing how well one set of basis images images can be used to reconstruct images from other sections of the data set are also tested on eyes open data. Figure 3.4.1.10 shows the first of these tests, and when compared to the results obtained on the eyes closed data, shown in Figure 3.4.1.8, are similar. Again, for analysis lengths less than 10 seconds the power represented by the first and second basis images vary considerably indicating less spatial variation in shorter intervals, and for lengths greater than 10 seconds the values are quite stable. In applying the basis images computed from the first 10 seconds of the record to other 10 second segments, shown in Figure 3.4.1.11, shows that the order of importance of the basis images is maintained. Table 3.4.1.3 demonstrates that the average reconstruction correlations are high, albeit lower than with the EC data.

The results from this section shed light on three basic questions regarding the spatial EEG. The first is that its intrinsic spatial dimensionality is lower than the measurement space, indicating that basic spatial patterns can be used to linearly reconstruct the spatial maps. The number of basis images required to adequately reconstruct the original background EEG data numbers no more than 5, with the underlying patterns discovered for the subject studied exhibiting back-front, front-back and side-side activity. The second point is that these basis images capture spatial patterns common throughout the record, as basis images from one segment can be used to represent, in a nearly optimal manner, images from other sections of the record separated as far apart as one minute. Furthermore, using 10 seconds of data to compute these basis images captures the underlying patterns and provides near optimal power representation for lengths of data ranging from 0.5 seconds to 40 seconds. Finally, there appears to be less spatial variation present in shorter intervals of the background EEG than in longer intervals, since power is more highly concentrated in fewer basis images for short data lengths.

3.4.3 Multiple Subject Study - Methods & Results

In order to verify that the claims regarding the spatial EEG made in Sections 3.4.1 and 3.4.2 are universal, similar analyses are required on other subjects. For brevity the detail used previously was not repeated, but rather analyses pertinent to the three questions asked in Section 3.4.1 were performed on several new subjects. The data from these subjects were acquired as described in Section 3.1.

For each of the subjects used in this section, denoted by Subject 2 EC, Subject 3 EC, etc., two sets of analysis were performed. (For brevity only EC data is used here). Figure 3.4.3.1 shows 64x64 pixel bicubic spline interpolated topographic representations of the first three basis images computed from the first 10 seconds of the data record. Figure 3.4.3.2 shows the effects on the power represented in each basis image for analysis lengths ranging from 0.5 to 40 seconds (similar to Figure 3.4.1.8a). Finally, Figure 3.4.3.3 gives the power represented by basis images 1-5 taken from the first 10 second segment of the record for other 10 seconds segments from the same record (similar to Figure 3.4.1.9), again for each subject.

For each of the ten projections taken per subject in Figure 3.4.3.3 the ARC was measured, with the average, maximum, and minimum of these values summarized in Table 3.4.3.1. (Recall that ARC values are based on the 31 point raw images only, not on the 64x64 interpolated images).

For all of the above analyses each electrode value from the original data was made zero mean for the analysis period.

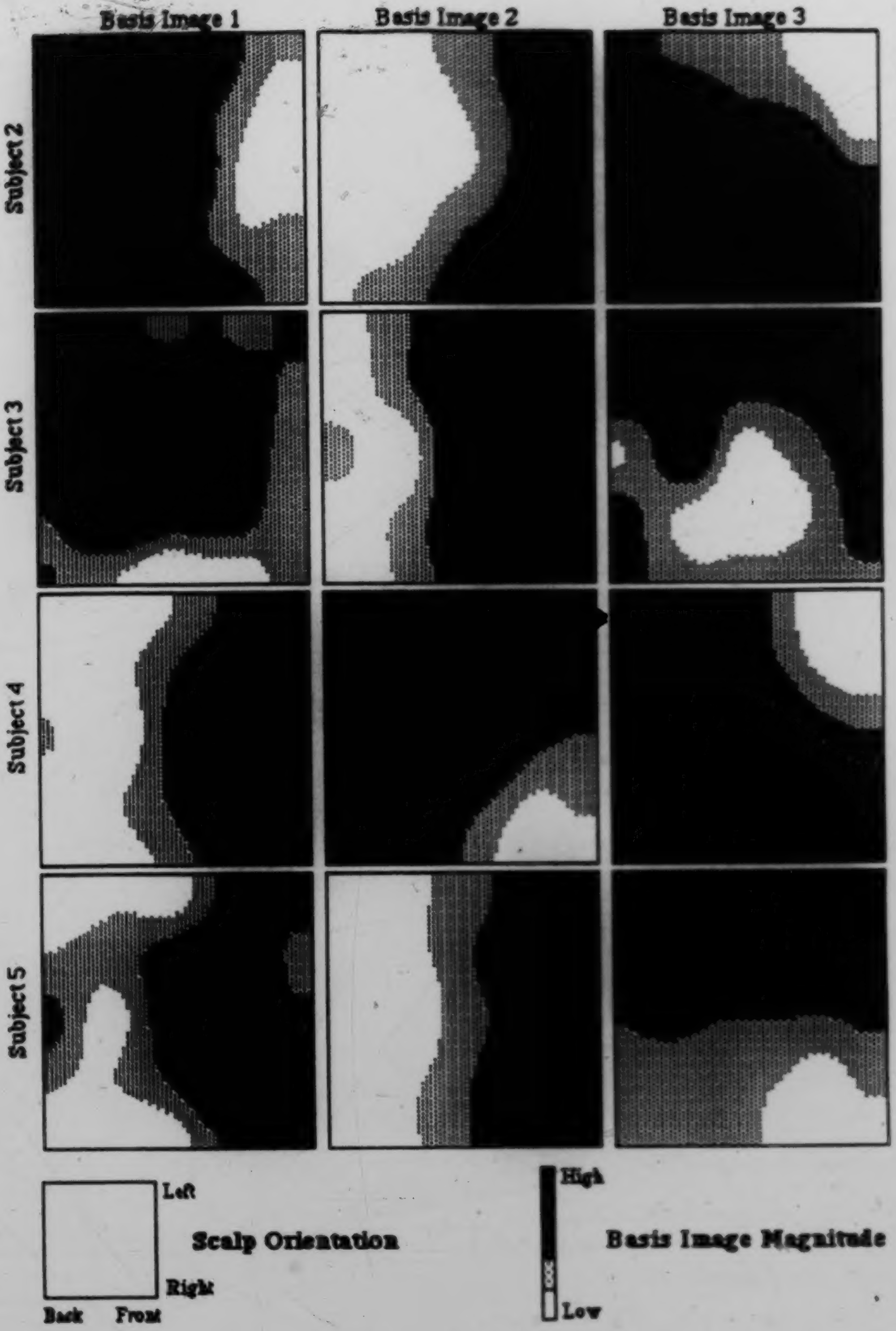


Figure 3.4.3.1
Basis Images 1-3 For Subjects 2,3,4,5 EC

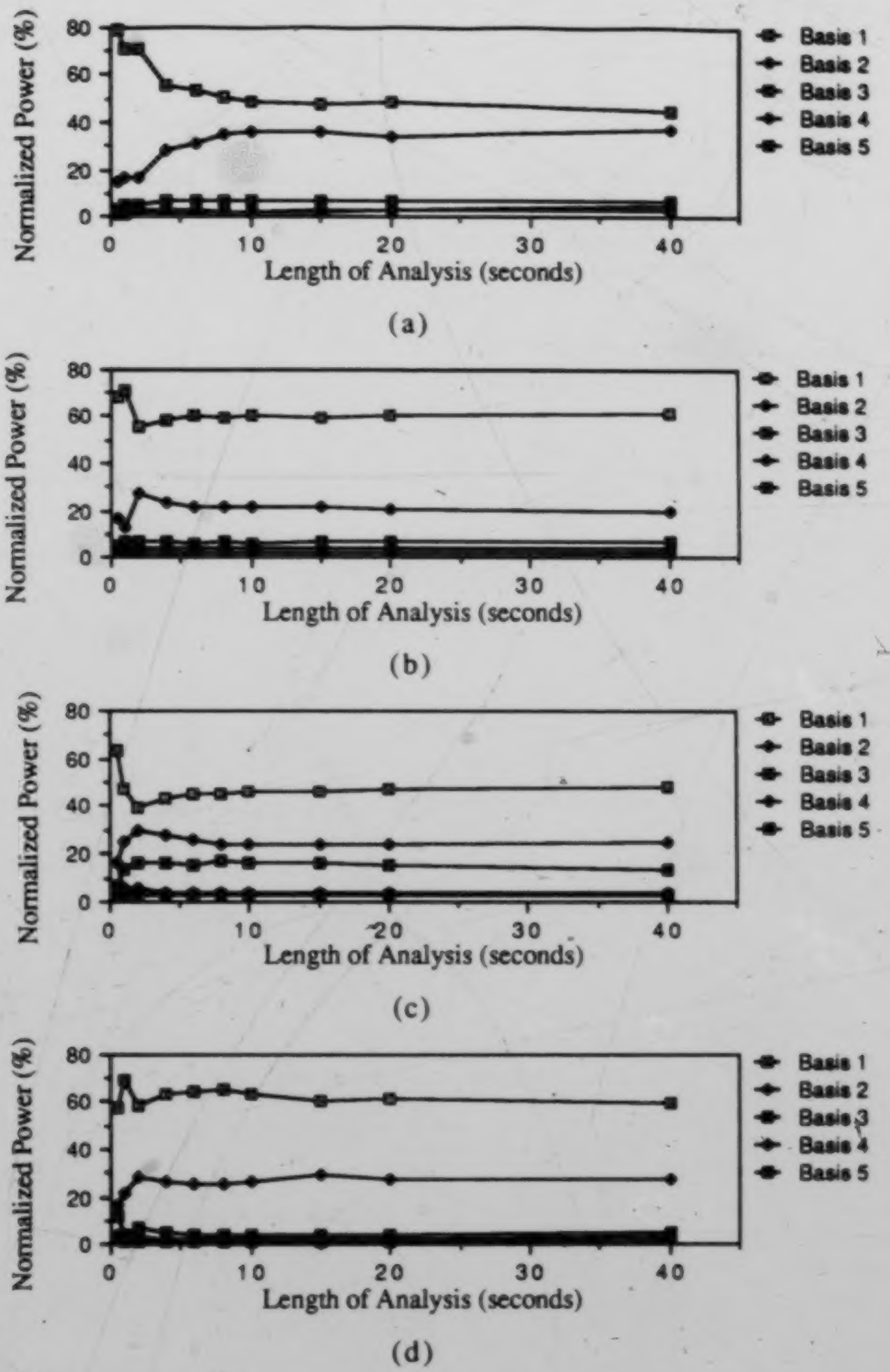


Figure 3.4.3.2

Power Represented by Optimal Basis Images For Varying Lengths of Data - (a) Subject 2 EC (b) Subject 3 EC (c) Subject 4 EC (d) Subject 5 EC

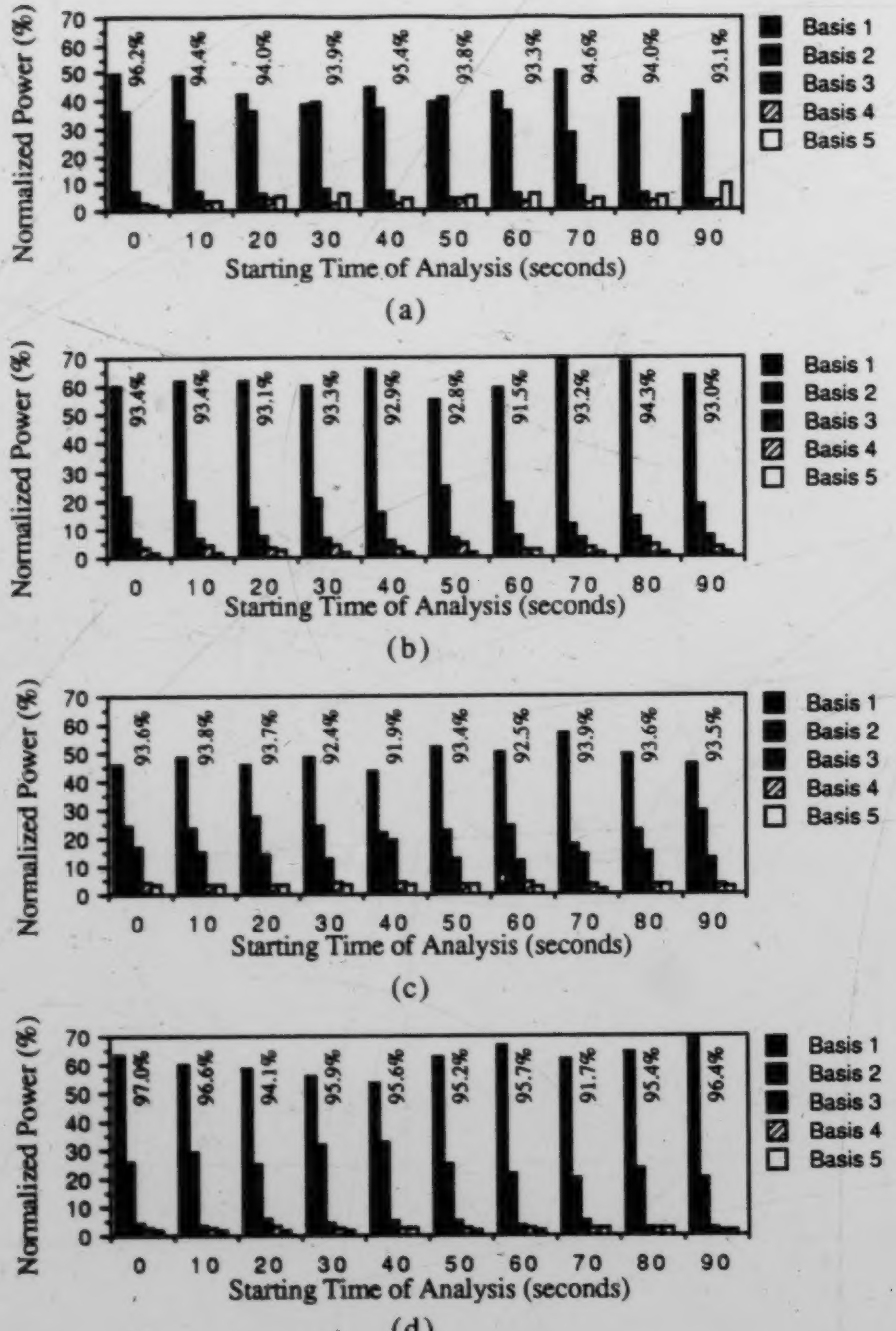


Figure 3.4.3.3

Power Represented by Basis Images Computed From the First 10 Seconds For 10 Second Segments of the EEG Record (a) Subject 2 EC (b) Subject 3 EC (c) Subject 4 EC (d) Subject 5 EC

	Average Reconstruction Correlation		
	Maximum	Minimum	Average
Subject 2	0.95	0.92	0.93
Subject 3	0.92	0.90	0.90
Subject 4	0.93	0.90	0.92
Subject 5	0.94	0.89	0.92

Table 3.4.3.1
Summary of Average, Maximum, and Minimum ARC Values Corresponding to Figure 3.4.3.3

3.4.4 Multiple Subject Study - Discussion

While there is a large amount of data displayed in Section 3.4.3 several patterns can be identified. One of the most consistent and important results, shown in Figure 3.4.3.2, is that for each subject, over 90% of the average power in the original images is represented in the first 5 basis images (regardless of the length of analysis). This indicates that for healthy individuals in the EC state the spatial EEG has an intrinsic dimensionality of around 5, reinforcing the idea that data compression can effectively and universally be performed.

Another key result is that when 8 or more seconds of data are used in computing the basis images the power represented by each basis image remains relatively stable. This indicates that using 10 seconds of data in the KLT analysis adequately captures the second order statistics of the spatial EEG, and that these statistics do not change significantly over time within the same cognitive state. As with Subject 1, the high concentration of image power represented by basis image 1 for short analysis intervals, (i.e. less than 1 second) reinforces the idea that less spatial variation is present in short segments of the eyes closed EEG.

Moreover, basis images computed using 10 second segments of data taken from an individual within one cognitive state can be used to represent other segments of data from the same state, separated as far apart as one minute, as shown in Figure 3.4.3.3. Reconstructed images using these basis images are similar to the original images, in terms of both power, and correlation as indicated in Table 3.4.3.1. This is useful for lowering the number of computations required for data compression, since an entire data record is not needed to estimate the basis images, but rather only a small section.

The basis images, shown in Figure 3.4.3.1, while exhibiting similarities, are unique to every subject. The most commonly occurring pattern indicated is front/back activity, which shows up for all individuals, typically as the first or second basis image. Another such pattern is a side/side activity, which again all subjects demonstrate in one of the three images shown, typically in the third basis image.

Several complex spatial patterns specific to individuals are exhibited, though. Basis image 1 of Subject 3 shows an area of high activity in the shape of an upside-down "U" extending from the center-rear to the center of the scalp. The third basis image from the same subject again has a complex pattern, with a valley located in the right-center of the scalp, and high activity shown on the left-center of the scalp. The second basis image of Subject 4 exhibits a diagonal topology. Basis image 1 of Subject 5 is wave-like pattern with low activity in the rear of the head followed by higher activity in the form of a longitudinal peak located just in from the front of the head.

In general, while common topologies of activity do exist, the details of these maps are unique to every individual. Furthermore the relative amounts of each basis image present in the data varies amongst the subjects. For example in Subject 2, as shown in Figure 3.4.3.3, the relative proportion of basis image 1 and 2 is quite close. In fact, the power represented by basis image 2 exceeds that of image 1 during 10 second epochs at 30, 50, 80 and 90 seconds into the record. For Subjects 3-5 basis image 1 represents a significantly larger amount of power than the other basis images, usually by over a factor of 2. Thus, while basis images computed from one section of data may be used to represent data from another section, the relative amounts of basis images vary considerably among subjects. As a result of these variations, using the proportion of power represented by each basis image as a method to classify the subject's cognitive state is not feasible.

The results from this section indicate that the results obtained from the detailed study performed in Section 3.4.1 are, in general, valid. Specifically: the intrinsic dimensionality of the spatial EEG for subjects in the EC state is lower than the measurement space, and is around 5; data compression using the KLT can significantly reduce the amount of storage required to represent the spatial EEG with the reconstructed images matching the original images in terms of both power and correlation; basis images computed from one section of a data record from the EC state can represent other segments of the same record, separated by as far as 90 seconds; less spatial variation is present in short segments of data than in lengthier segments; and finally, approximately 10 seconds of data adequately captures the spatial properties of the EEG for lengths ranging from 0.5 to 40 seconds.

3.5 Pattern Representation of the EEG Using the Karhunen-Loeve Transform on Subject Groups

Section 3.4 showed that the intrinsic dimensionality of the spatial EEG *within an individual in a particular cognitive state* is lower than the measurement space; that is, raw EEG data can effectively be represented by the linear combination of basis images. Using these basis images results in good reconstruction accuracy in two senses: most of the power (variance) in the original images can be captured in several basis images, and the reconstructed images correlate, on average, highly with the original images.

In this section the second of the three questions posed in Section 3.0 is investigated, more specifically, are the basis images and coefficients similar between different cognitive states in the same individual, and from one subject to those of other subjects? The answer to the latter question is useful in speculating whether the basis images are physiologically meaningful, since it is expected they be similar for subjects in the same cognitive state. This of course assumes that the same basic physiology is responsible for similar mental processes amongst individuals, and that individual variations of skull thickness, scalp conductivity and other such factors are negligible. Otherwise the KLT may be extracting information highly dependent upon the individual, limiting the usefulness of the KLT as a tool to understand the EEG (since there would be no common framework in which to compare and classify individual data).

To answer these questions three routes of investigation are taken. The first tests whether the basis images from a subject in one cognitive state adequately represent spatial maps from the same individual in another cognitive state. Several possibilities exist from this analysis: if the above test is true and the power distribution per basis image is similar then the KLT is suggesting that no significant difference in the second order statistics of the spatial EEG from each cognitive state exists. If the above test is true and the power distribution is different, then each cognitive state may be thought of as different mixture of the same underlying spatial distributions, and thus the coefficients may be used to classify the cognitive state *within an individual.* If the above test proves false then the cognitive states tested have unique spatial voltage patterns associated with them and pattern representation must be performed using data from within a cognitive state only.

The second route of investigation checks if basis images from one individual can represent spatial images from another individual in the same cognitive state. This is akin to checking how similar the basis images from one subject are to those from another subject. If such a test is true then, indeed, there exist common patterns of spatial activity amongst the individuals tested, and therefore perhaps in the population in general.

The final test checks whether a set of basis images computed from a concatenation of the five subjects' data can represent the original data from those individuals. This is really a refinement upon the previous analysis, except that the basis images should provide a more universal basis set since they are taken from the common characteristics of all subjects.

As before Equation 3.4.1 is used to compare the normalized percentage of power represented by each basis image. Since power in the EEG varies considerably between individuals, the usefulness of using such a normalized value should now be clear as only representation ability, independent of signal power, is measured. For brevity, when the term "power" is referred to in the following sections the normalized measure of average image power is intended.

The analyses performed in the following sections was necessarily limited, as there were many possible combinations of tests even for the five subjects studied in Section 3.4. Although the investigation was not extensive, the results do indicate areas for further research.

3.5.1 Pattern Representation Across Cognitive States Within and Individual - Methods and Results

In this section basis images computed from the EC state were used to see how well they represent data obtained from the same individual during the same recording session, in the EP state. As in Section 3.4 basis images computed from 10 seconds of images were used, as they adequately capture the properties of the spatial EEG over lengthier periods. As before, each electrode sequence of the data being studied was made zero mean for the analysis period to remove the effects of amplifier drift, e.t.c.

In this section data from each of the five subjects investigated in Section 3.4 was used. The bar graphs in Figure 3.5.1.1 show the distribution of basis function power when 10 second segments of EP data were projected on to basis images 1-5 computed from the first 10 seconds of the subject's EC record (refer to the appropriate figures in Sections 3.4.1 and 3.4.3 for a topographic representation of the basis images). The numbers above each group give the total power.

Table 3.5.1.1 summarizes the average, maximum and minimum ARC values over each the 6 projections in Figure 3.5.1.1 for each subject.

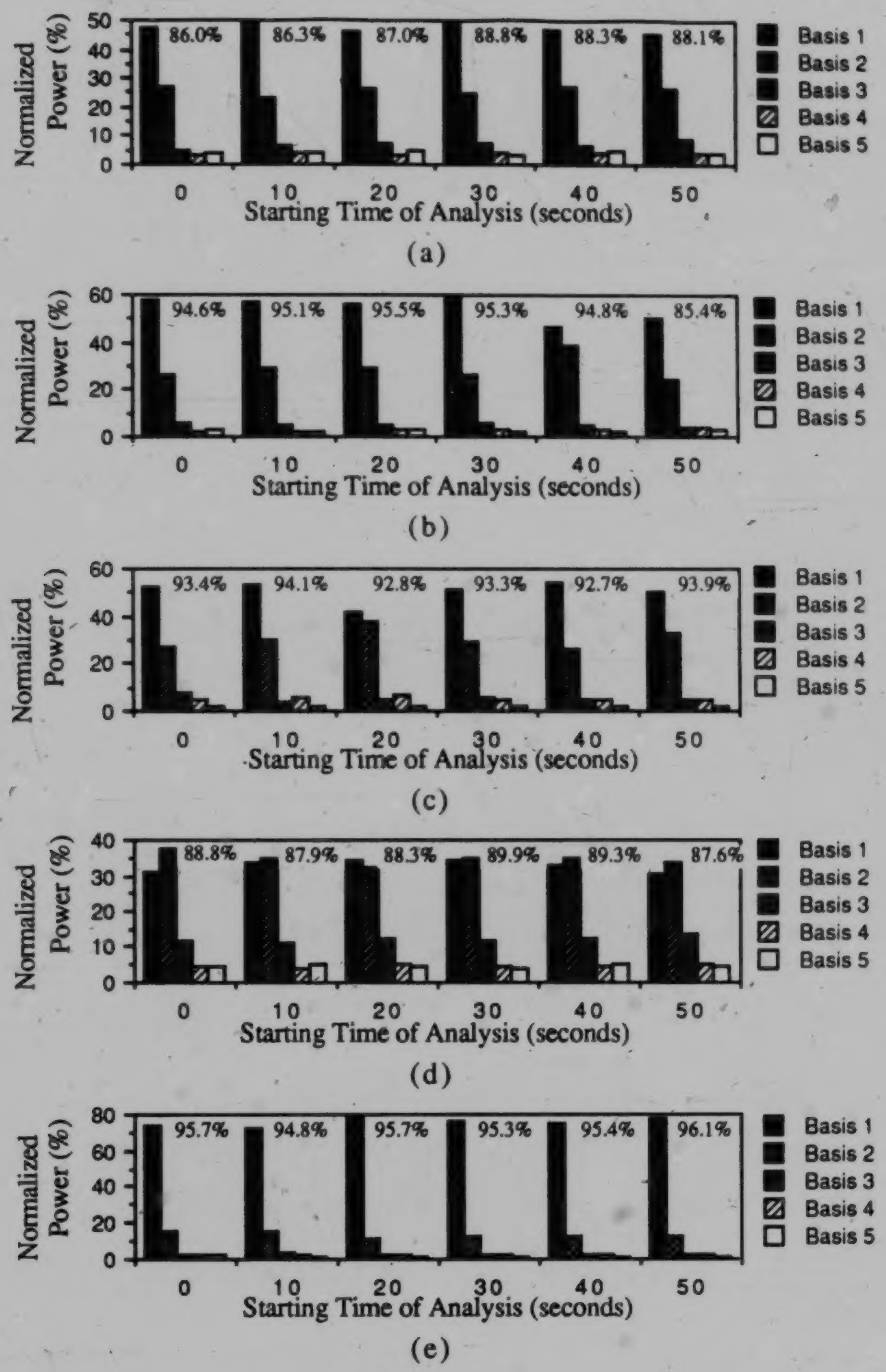


Figure 3.5.1.1

Power Represented by EC Computed Basis Images for 10 Second Segments of EP Data - (a) Subject 1 (b) Subject 2 (c) Subject 3 (d) Subject 4 (e) Subject 5

	Average Reconstruction Correlation		
	Maximum	Minimum	Average
Subject 1	0.87	0.83	0.85
Subject 2	0.93	0.92	0.93
Subject 3	0.92	0.91	0.91
Subject 4	0.89	0.87	0.88
Subject 5	0.92	0.91	0.92

Table 3.5.1.1

Average, Maximum, and Minimum ARC Values For Figure 3.5.1.1

3.5.2 Pattern Representation Across Cognitive States Within an Individual - Discussion

Table 3.5.1.1 and Figure 3.5.1.1 show that both power and reconstruction accuracy are generally high when basis images from one state are used to represent images from another state within the same individual. However, the results are not as good as when EC data is projected on to EC basis functions. The above bar graphs compared to the corresponding ones for EC data (Figure 3.4.1.9 for Subject 1, and Figure 3.4.3.3 for Subjects 2-5) exhibit that the percentage of power represented by each basis image varies between cognitive states.

For example, in Subject 1 when EC data and basis images are used the normalized average power represented by the first five basis images is always over 90%, with the ARC values, shown in Table 3.4.1.2, typically over 0.90 (average of 0.92). Figure 3.5.1.1a shows that those same basis images represent less than 90% (average of 87.42%) of the power in the EP data examined, and the corresponding

ARC values average only 0.85. Recall that ARC values are significant to within +/-0.01, and that power values are significant to +/-1%, both using a 99% confidence interval.

The next question is how do the EC basis images compare to those computed from the EP state, in terms of representing EP data? Figure 3.4.1.11 and Table 3.4.1.3 show EP computed basis images fitted to the same EP data as above. Here the first five basis images represent over 90% of the power in the segments analyzed, with the ARC values averaging 0.89. For this subject therefore, EC basis images do not represent EP data as well as EP basis images.

The results for Subject 2 indicate for this subject that EP data is represented by EC basis functions with high accuracy. When the five EC basis images are fitted to EC data they correlate on average 0.93, as shown in Table 3.4.3.1, and when fit to EP data correlate, as shown in Table 3.4.5.1, on average 0.93, with over 90% of the power typically represented. Although results for the optimal case have not been computed for this subject, the basis images and therefore data in the two cognitive states must be similar, based upon the high accuracy found above. The power represented by the basis images varies between each state, though, as shown by comparing the appropriate bar graphs in Figures 3.4.3.3 and 3.5.1.1. In the EC state the amount of basis image 1 and 2 required is similar, while for the EP state basis image 1 increases and 2 decreases in relation to the EC results.

The results from Subject 3 indicate good representational ability: for power averaging 93.4%, and ARC on average 0.91, compared with 0.92 for the EC data. The distribution of basis function power in the EP state shows an opposite trend to that of Subject 2. More specifically, the proportion of basis image 1 drops, and of basis image 2 increases in relation to the EC data. These results again indicate that the underlying patterns in the EP state are similar to those of the EC state, with the amount of the each basis pattern present varying between states.

The reconstruction ability using EC basis images on EP data for Subject 4 is poorer than that for EC data with the same EC basis images. The power represented by the first 5 basis images drops to less than 90% for each segment examined, and the ARC is 0.88, compared to 0.92 for the EC data. The proportion of power between basis images 1 and 2 changes dramatically between the EC and EP states, with basis image 1 representing significantly more power (typically over 25%) than that of basis image 2 for the EC state, and both representing approximately the same power for the EP state.

The representational ability of the five EC computed basis images for Subject 5 EP data is high, both in terms of power, well over 90%, and shape, correlating on average 0.92. Here the power represented by basis image 1 increases, and for basis image 2 decreases, in the EP state.

In general, EC computed basis images from one subject can be used to represent EP data from the same subject, although the reconstruction ability is suboptimal both in terms of the percentage of average power represented, and reconstruction correlation. Therefore, for data compression it is advisable that basis images computed from one cognitive state be used to represent data only from within that same state.

In differentiating between the two states by comparing the distribution of power represented by the basis images there appears to be no single trend amongst the individuals studied. This may indicate one of three things: that tighter experimental controls are required to ensure that, in fact, all subjects are in the "same" cognitive state; that a finer division of subjects according to sex, age, e.t.c is required to elicit any trends; or that the difference between these two cognitive states is dependent upon the individual. It is worth noting that while there is no single trend between EC and EP data, there are detectable differences in the distribution of power represented by the basis images. Using the KLT directly, it therefore may not be possible to determine the actual cognitive state the subject is in, but it may be possible to detect a *change* of state within an individual.

3.5.3 Pattern Representation Across Individuals in the Same State - Methods and Results

In this section basis images computed from an individual in one cognitive state were used to represent data from another individual in the "same" cognitive state. This, in essence, measures how similar basis images from one subject are to those of another. As there are many possible combinations of investigation for the five subjects only a subset was performed.

Figure 3.5.3.1 shows the distribution of power when 10 second segments of data from the EC records of Subjects 2 through 5 are projected on to the first five basis images, computed from the first 10 seconds, of Subject 1's EC record. Table 3.5.3.1 shows the average, maximum, and minimum ARC values for each of these projections. The basis images used are those displayed in Section 3.4.1.

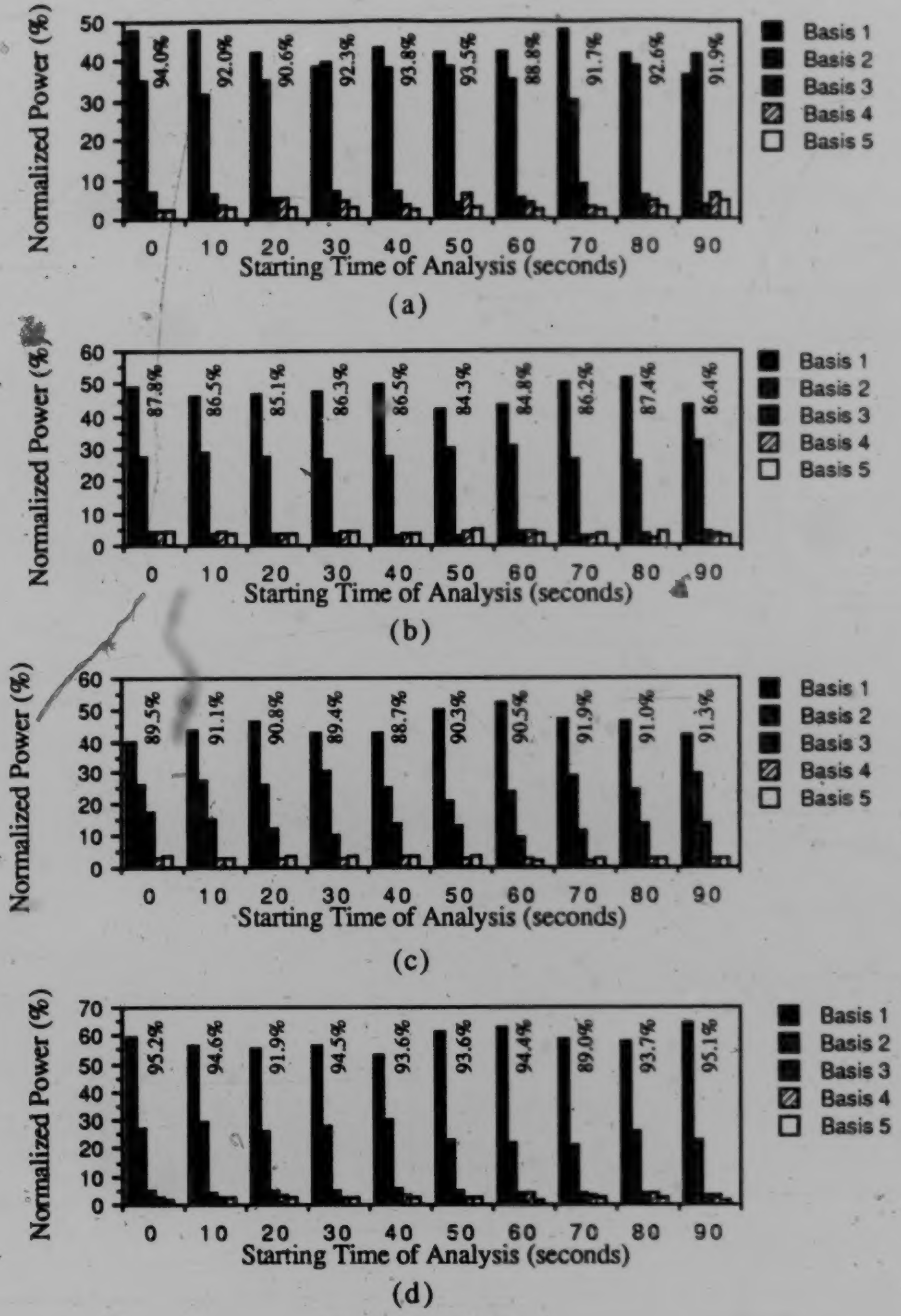


Figure 3.5.3.1

Power Represented by Subject 1 EC Basis Images for 10 Second Segments of EC Data (a) Subject 2 (b) Subject 3 (c) Subject 4 (d) Subject 5

	Average Reconstruction Correlation		
	Maximum	Minimum	Average
Subject 2	0.92	0.88	0.91
Subject 3	0.85	0.81	0.82
Subject 4	0.90	0.87	0.89
Subject 5	0.92	0.85	0.90

Table 3.5.3.1
Average, Maximum, and Minimum ARC Values For Figure 3.5.3.1

3.5.4 Pattern Representation Across Individuals in the Same State - Discussion

The results indicate that, for the cases investigated, basis images from one subject can be used to represent data from other subjects in the same cognitive state. The representation is not as good, though, as using basis images computed from the same state and individual.

One major trend appears in the results, confirming an obvious suspicion, and that is when the basis images between the two subjects are similar in appearance the reconstruction accuracy is high, while, if the basis images are dissimilar the accuracy is low. For example, the basis images between Subject 1 and 2, shown in Figures 3.4.1.1 and 3.4.3.1 respectively, are similar. Both the reconstruction correlation, 0.91, displayed in Table 3.5.3.1, and average power represented by the basis images, shown in Figure 3.5.3.1, are similar to those values obtained using Subject 2's basis images, shown in Table 3.4.3.1 (0.93) and Figure 3.4.3.3 respectively. The results for Subjects 4 and 5 are similar to those for Subject 2.

On the other hand, Subject 3's basis images, shown in Figure

3.4.3.4 differ significantly in appearance to those of Subject 1. The average reconstruction correlation using Subject 1's basis images is 0.83, compared to 0.90 using Subject 3's own basis images. The distribution of power is also different, with basis image 2 increasing in percentage, basis image 1 decreasing, and the total using the first five dropping to an average of 86.1% from 93.2%.

These results indicate that representing a subject's EEG with basis images computed from another individual is not generally as effective as using the subject's own data, and is related to the similarity of the basis images themselves. It is interesting to note that there exists more similarity in the basis images amongst individuals in the same cognitive state, than from within an individual in different cognitive states.

Again, no specific trends are seen in the results. These trends may, for example, indicate that a subject's basis images are useful for representing data only from individuals in the same age group, sex, handedness, e.t.c. To detect such correlations more subjects and categorization of them along the above lines is required.

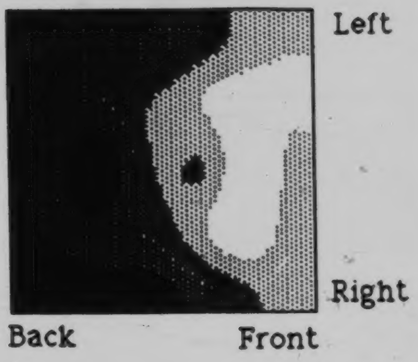
3.5.5 Pattern Representation Using Collective Data - Methods and Results

In this section basis images are generalized by combining the data of several individuals before KLT analysis. This is performed in order to capture the common spatial patterns amongst the individuals studied. If such patterns do exist then the patterns detected may also be indicative of the population in general.

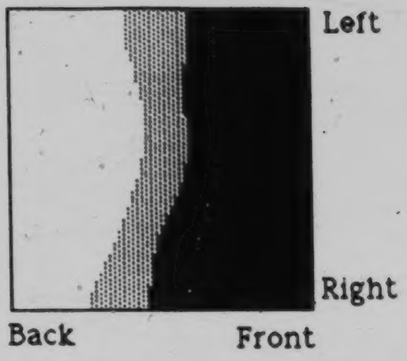
One problem in using data from various subjects is that some individuals have more power in their EEG than others. Since fitting to these subjects data would minimize the mean squared error first, their data would easily dominate the KLT analysis, and the computed basis images would highly resemble their own. In order to avoid this problem the data was prenormalized before applying the KLT to capture shape, not magnitude, information. The method employed normalized each image such that the sum of squares of all pixel values, (electrode samples for "raw" data, as used here) is unity. Before applying this normalization each measurement value *within* each subjects data segment was, of course, made zero mean. This ensured that the mean values subtracted from the electrode recordings corresponded to the recording session and individual data only, not from the collective set of data.

The data used in this section was a 50 second record, created from the first ten second segments of each of the five subjects previously studied. As stated above, each measurement value within each the five segments was made zero mean, and then each image (i.e. 31 measurement values) was normalized to have a sum of squares equal to unity.

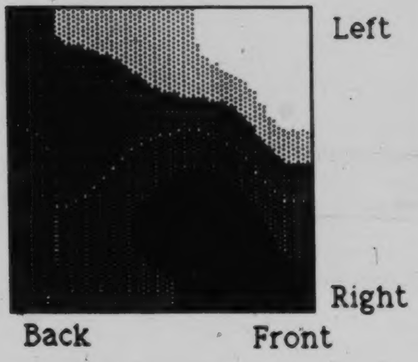
Figure 3.5.5.1 shows topographic representations of the first five basis images computed from the 50 second record, while Table 3.5.5.1 shows the normalized power represented by the first ten basis images. Figure 3.5.5.2 displays the normalized power represented by each of these basis images when 10 second segments of data, taken from each of the five subjects, are projected on to the collective basis images. Finally, Table 3.5.5.2 shows the average, maximum, and minimum ARC values of the 10 projections performed on each subject's data.



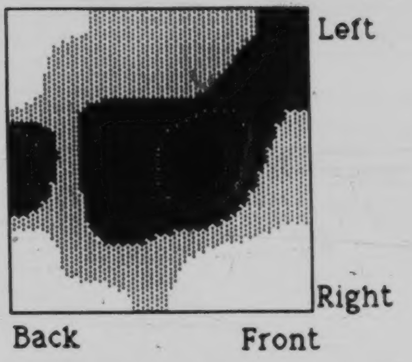
Basis Image 1



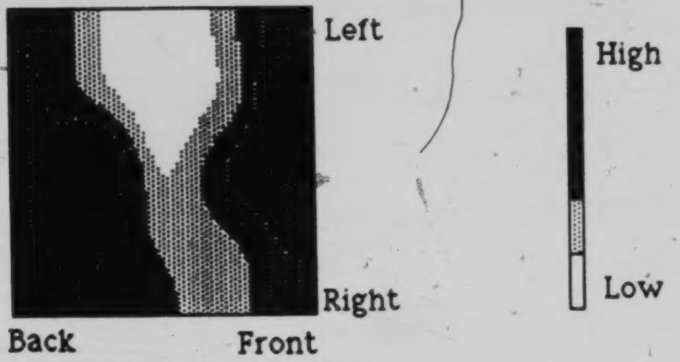
Basis Image 2



Basis Image 3



Basis Image 4



Basis Image 5

Figure 3.5.5.1

Basis Images 1 - 5 For Collective Data Set

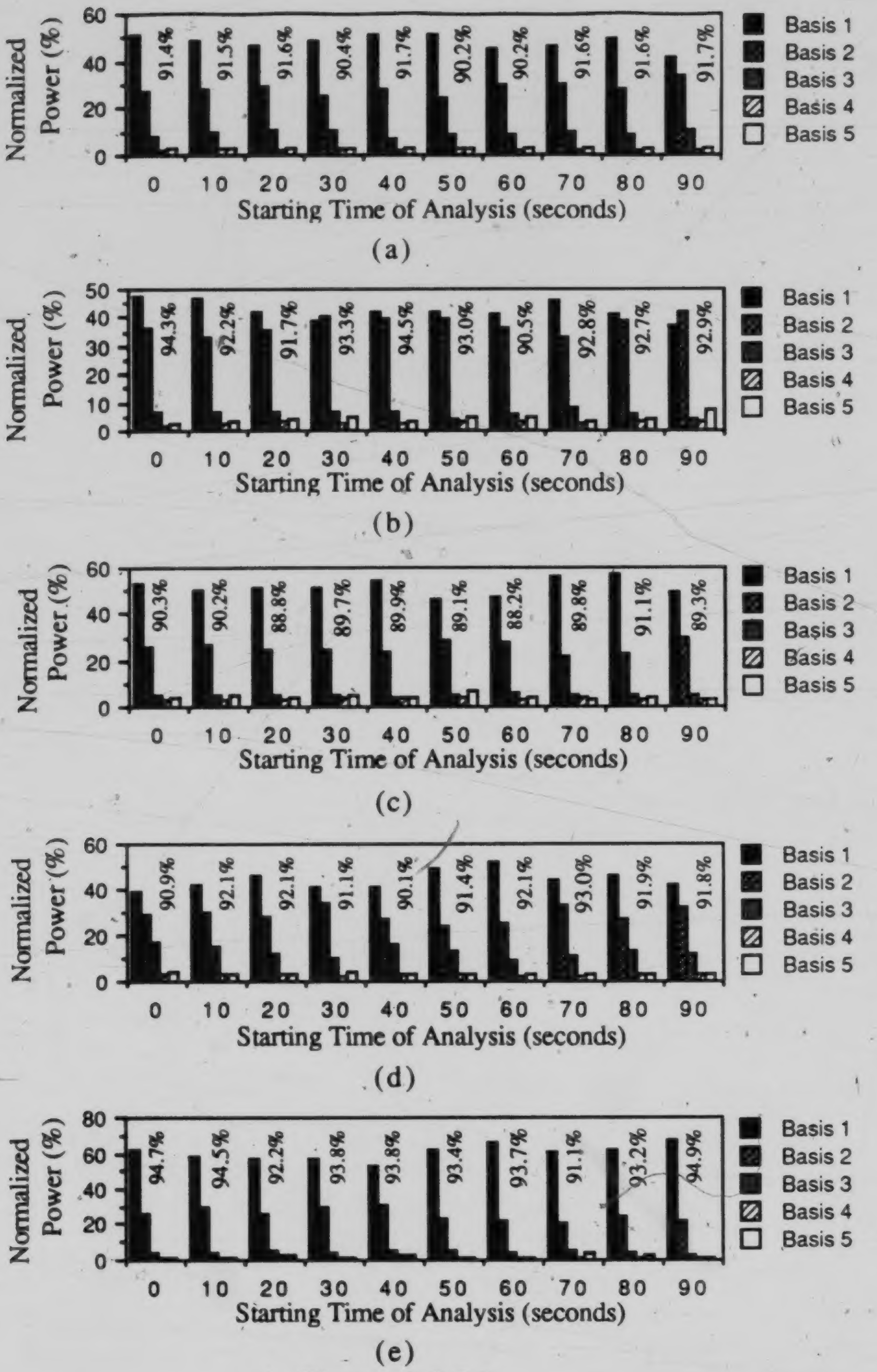


Figure 3.5.5.2

10 Second Segments of Data Projected on to Collective Basis Images - (a) Subject 1 (b) Subject 2 (c) Subject 3 (d) Subject 4 (e) Subject 5 EC

Basis Image	Normalized Power (%)	Cumulative Power (%)
1	39.98	39.98
2	26.48	66.46
3	10.14	76.60
4	7.21	83.81
5	3.76	87.57
6	2.89	90.46
7	2.49	92.95
8	1.42	94.37
9	1.03	95.40
10	0.77	96.17

Table 3.5.5.1
Normalized Power For the First 10 Basis Images Computed from the 50 Second Collective Data Segment

	Average Reconstruction Correlation		
	Maximum	Minimum	Average
Subject 1	0.89	0.87	0.88
Subject 2	0.93	0.89	0.91
Subject 3	0.88	0.85	0.87
Subject 4	0.91	0.89	0.90
Subject 5	0.91	0.90	0.90

Table 3.5.5.2
Average, Maximum and Minimum ARC Values For Figure 3.5.5.2

3.5.6 Pattern Representation Using Collective Data - Discussion

The above results demonstrate that the intrinsic spatial dimensionality of the collective data set is, as would be expected, higher than individual data. Although this is the case, the reconstruction ability of the collective basis images for data from the same cognitive state is good, and is better than using data from a single subject to represent data from other subjects.

Table 3.5.5.1 shows that more basis images are required to represent the same power for the collective data set, than for each of the subjects individually. This indicates that more spatial variation is present in the EC state amongst individuals than within a subject.

The basis images, shown in Figure 3.5.5.1, exhibit similar features to those of the individuals used to create the collective data set, displayed in Sections 3.4.1 and 3.4.3. The first and second basis image demonstrate front/back activity, although the first has a more complex pattern, with what appears to be a "hot" spot located over one of the central electrodes. Basis image 3 is similar to the third basis images from Subjects 1, 2, 4 and 5 (if the polarity for Subject 5 is changed), containing an area of high activity in the front right, and low activity in the front left of the scalp. Basis images 4 and 5, not shown for the individuals, have unique spatial patterns: image 4 has high activity along the mid-line of the scalp with two peaks of activity, centered on electrode locations; image 5 is wave-like with high activity in the front and rear of the scalp and low activity in the middle.

The bar graphs in Figure 3.5.5.2 show the distribution of power represented by the collective basis images for Subjects 1-5, with Table 3.5.5.2 giving the average, maximum, and minimum ARC values for each of the 10 projections performed per subject. For the analyses performed in the previous sections the best results for such projections have been obtained using basis images computed from the data itself (optimal case), followed by using data from one segment to represent other sections. The third best results were obtained using basis images computed from another individual in the same cognitive state, with the poorest all around performance obtained using basis images computed from the same individual in another cognitive state.

The above results compared with the those obtained using the subjects own data do not provide as good an overall representation. For example, for Subject 1, the power represented by the first five basis images closely matches that using Subject 1's own basis images, shown in Figure 3.4.1.9(b), with over 90% of the average image power represented in the first five basis images in both cases. The average, over the 10 projections, of the ARC values, is 0.92 using the subject's own basis images, and 0.88 using the collective basis images. Thus, average image power for the former case is represented as well as for the latter, while the shape is not.

For Subject 2, the power represented by the first five collective basis images again closely matches that using the subject's own basis images, displayed in Figure 3.4.3.3. The average ARC using the collective basis images, 0.91, is lower than using the subject's own basis images, 0.93, and the same as using Subject 1's basis images, 0.91.

The power represented by the collective basis images for Subject 3 is less than that obtained by using the subject's own basis images, but is better than using those from Subject 1. This, again, is probably due to the significant difference in appearance between the basis images from Subject 3 and those from the collective data. This difference is also reflected in an average reconstruction correlation of 0.90 using Subject 3's data, and 0.87 from the collective data. This latter value is an improvement, though, over using Subject 1's basis images (0.83).

For Subject 4 the power represented using the first five collective basis images is typically over 90%, although the distribution amongst the basis images is different from that of using the subject's own basis functions (basis image 1 has decreased, and 2 increased in power). The average ARC is slightly lower than using the subject's basis images, 0.90 compared with 0.92, and the same as the value obtained using Subject 1's basis images.

The results from Subject 5 show similar trends to those of Subjects 1 and 2, that is, the distribution of power is almost the same using the collective basis images, with the corresponding reconstruction correlation lower than using the subject's basis images (0.89 compared with 0.92). The only difference is that the average ARC value using Subject 1's basis images, 0.90, was slightly higher than from the collective data.

In general, the above results indicate that using basis images computed from collective data rather than from an isolated individual is useful in representing each of the subjects contributing to the collective data. The results also indicate that these basis images provide better representational accuracy than using a single individual's data. The collective basis images exhibit properties present in all of the subjects' basis images, again front/back activity and side to side activity appear to be the most prominent spatial patterns. These images can therefore be regarded as spatial features common to all individuals in the cognitive state investigated. Further study is required to determine if such patterns are common throughout the entire population; what, if any, physiological basis such patterns have; and if the basis images and coefficients can be used to classify between cognitive states and individuals.

3.6 Discussion and Conclusions

In this chapter the Karhunen-Loeve transform has been used as a tool for pattern representation of the spatial EEG. Several general conclusions can be reached, while many avenues have yet to be pursued in order to determine the exact effects that cognitive state, age, handedness (Gevins, 1987), e.t.c have upon the products of KLT analysis. The results from this chapter confirm observations seen by other groups which have applied the KLT to EEG analysis.

The main result found is that the intrinsic dimensionality of the spatial EEG is less than the measurement space used for the observations. More specifically, usually over 90% of the average power in spatial images is represented by typically using the first 5 basis images (and certainly no more than 10), independent of the cognitive states investigated. Images reconstructed using these basis images and the appropriate coefficients also correlate, on average, highly (typically 0.90) with the original data images.

The spatial dimension of temporal sequences of background EEG being lower than the measurement space implies that data compression can be performed on raw EEG data with significant savings in on-line storage. For example, storing the first 20 basis images (more than necessary for adequate reconstruction) computed from minutes of recorded data requires approximately 20/31, or about 65% of the original storage requirements. If the spatial basis patterns are physiologically meaningful, then this intrinsic dimensionality should be independent of the measurement space, and thus greater savings will be possible for larger numbers of electrodes, with the added benefit that the basis images will have better resolution because of the increased spatial sampling.

An important result, useful in data compression, and significant to the EEG, is that basis images computed from one section of an individual's record taken from one cognitive state can be used to represent, in a nearly optimal manner, data from other segments of the same record. Thus the basic spatial patterns generating the maps are not changing significantly over time within one cognitive state, indicating that an underlying process may be generating the spatial topologies seen. Analysis of varying length data demonstrates that for short segments (<1 second) little spatial variation exists, while for lengths greater than about 8 seconds the spatial variation is larger and independent of length.

A point related to the above, is that the first several basis images computed from one state can be used to adequately represent data from another state, although the proportion of the coefficients varies between states. This variation, for the subjects investigated, while being detectable shows no evidence of a uniform trend. Again, breakdown of data along finer subject groups would help illuminate any trends.

It was also shown that basis images between individuals are similar, in a pattern representation sense. While this is not clear evidence of a common process generating cognitive states in individuals it does provide a mechanism for investigating these processes. As above, finer division of subject groups is required to identify trends useful for classifying the data by cognitive state.

In general, although no clear classification rules are apparent from the results obtained here, the usefulness of the Karhunen-Loeve transform as a method to extract common features of the spatial EEG, both within and across subjects has been demonstrated. The ability of the KLT to generate features useful in classifying data is examined in the following chapter.

Although these results are for the background EEG taken under the eyes closed (EC), and eyes open (EP) states, they are in general agreement with similar studies performed on spatial maps from evoked potential studies, where dimensionality reduction has repeatedly been shown possible (Skrandies and Lehmann 1982; Chapman et al. 1979; Kavanagh et al. 1976; Donchin and Heffley 1978), and from other studies where both spatial and temporal properties of the EEG were used (Nunez 1981).

Skrandies and Lehmann (1982) found that the spatial KLT yielded three spatial components, accounting for over 93% of the variance in data obtained by retinal stimulation. The basis components were computed from images from six subjects at 100 and 140 msec latencies, where it had been estimated maximal global field power occurred. The first component basis exhibited back/front activity, the second lateral (side/side) activity, and the third a concentric distribution with high activity in the center-rear of the scalp. Skrandies and Lehmann demonstrate these component images are physiologically meaningful and in agreement with analyses performed on the evoked wave shapes. It must be remembered that their study did not analyze the temporal evolution of the scalp distributions, but rather performed the analysis across subjects.

An interesting point in their paper is that the second basis image computed at the 100 and 140 msec latencies were "mirror" images, and they suggested that this "might reflect the complete activity of one functional population of generators" (p. 666). The same trend is seen in the first two basis images of most of the subjects studied in this chapter and may reflect the operation of a similar process. Further investigation is required, although these results indicate that the KLT applied to the background EEG is extracting relevant physiological information. Another point worth mentioning is that the spatial dimensionality found by Skrandies and Lehmann is lower, around 3, compared to that for background EEG found here, about 5. This difference may be a reflection of the specificity of the tasks the subjects in their study were performing or be a product of the temporal character of analysis performed here.

Another study reported is by Nunez (1981), where he uses "empirical orthogonal functions in the spatial domain". In his analysis the spatial functions do not correspond directly to the background EEG, but rather to discrete temporal frequencies of the EEG, that is, to components of the Fourier spectrum. The basis functions are therefore complex and preserve phase information. This method has the advantage that spatial information is captured for each temporal frequency of interest, for example, alpha band activity. The obvious disadvantage is that there are several sets of basis images to consider, one set per temporal frequency of interest.

The results reported by Nunez are similar to those presented in this Chapter. For a patient with peak alpha activity at 9.5 cps Nunez gives the first three basis images (both the real and imaginary parts) corresponding to to 9.0, 9.5, and 10.0 cps for one minute of EEG (30, two second epochs). At the peak frequency the first basis image accounted for 92% of the power in the original data, with the first basis image for the off center frequencies representing 63% and 79% of the average power. Again front/back and side/side activity were present in all real basis images at each frequency, with the front/back topology predominating.

Nunez's results show that the shape of the basis images are frequency dependent, and that the highest data compression occurs at the peak temporal frequencies. Nunez also shows that the coefficient of the first basis image reflects the amount of alpha activity present temporally and can thus be used to differentiate between eyes open and eyes closed, although he does not provide a quantitative means for doing so.

While the research reported by Nunez is not exactly the same as what is performed here, the results he presents confirm those obtained here. Whether it is better to break down the spatial maps according to temporal frequencies or not is an open question. The relationship between temporal variation of the scalp potentials and the underlying physiological bases for the EEG is required before such an question can be answered.

In general then, this chapter has demonstrated the usefulness of the KLT as an efficient pattern representation, and hence data compression, technique for the background EEG. While no specific relationships between the basis images of different individuals has been found, the results suggest that similarities in inter-subject basis images taken from the same cognitive state may exist. The techniques and results from this chapter therefore provide the groundwork for further investigation using the KLT into the connection between the EEG and cognitive state.

Chapter 4

Pattern Classification of the EEG Using the Karhunen-Loeve Transform

4.0 Introduction

The last, and most difficult, application of the KLT to the EEG pursued here is to determine if the basis functions and/or coefficients can be used to discriminate between various cognitive states. In this sense the KLT is viewed as a technique which extracts features from a set of measurements made on a realization of a statistical process. These features can later be used to classify the data. Recall that feature extraction corresponds to the first basic step performed in statistical pattern recognition discussed in Chapter 2; in this context the KLT is popular and has been extensively used (Kittler 1977; Kittler 1974; Kittler 1986; Periyalwar et al. 1987; Van hamme et al. 1987; Shimura and Imai 1973; Kittler and Young 1973; Fukunaga 1972; Skrandies and Lehmann 1982; Kennett 1983; Fukunaga and Koontz 1970; Glaser and Ruchkin 1976; Schmeidl et al. 1987).

Many variations based upon the KLT are available for pattern classification, Kittler (1977) mentions eleven of the most popular techniques. Each of these methods takes advantage of different strengths in the KLT, with most applying a transformation to the data either before or after analysis. These transformations attempt to maximize the differences in pattern vectors by some preselected criterion since, in many cases, the fundamental components from two related pattern classes may be similar.

The transformation method applied in this thesis is the one proposed by Fukunaga and Koontz (1970), for the two class problem of classification. They do not derive "optimal" features for classification, but rather transform the measurement coordinate system before applying the KLT resulting in "good" features for classification. The results of this transformation yield basis functions which are the same for both pattern classes, but whose eigenvalues are reversely ordered. That is, the basis function whose eigenvalue is largest for one class will have the smallest eigenvalue for the other class. Classification of the data can then proceed by applying the transformation function to new data, performing KLT analysis, and noting the order of importance of the eigenvectors, or basis functions.

Another popular transformation which has been applied to image sequence analysis (Van hamme et al. 1987), and which for completeness should be mentioned here, is the method of Kittler and Young (1973). In their method emphasis is placed on using class mean information for optimal classification. Optimal features are extracted when the class mean values are useful for categorization, for example, when pixel sequences representing a tumour, etc., have a different mean intensity than other tissues. In sequences of raw EEG spatial maps all classes (pixel sequences) of data are zero mean, and for this reason their transformation cannot be used here. If, on the other hand, current density maps or power spectrum images were analyzed, the method of Kittler and Young might prove useful for image classification, recognition of tumour sites or lesions.

In the next section a review of the Fukunaga-Koontz transformation (FKT) is given, followed by the results obtained by using their method on spatial maps of the EEG. Section 4.2 demonstrates the classification ability of the FKT between individual subjects' EC and EP data. Section 4.4 shows the classification ability when a single subjects' FKT features are used on other subjects' data. Finally, a collective set of FKT features is used for classification in Section 4.6.

4.1 The Fukunaga-Koontz Transformation

The notation used here is consistent with Chapter 2, with any extensions following that of Fukunaga and Koontz. The development used here is based on the theory presented in their paper.

The Fukunaga-Koontz transformation (FKT), when applied to two known pattern classes, yields a set of common basis functions whose corresponding eigenvalues are reversely ordered. The method is based upon a class combined autocorrelation matrix defined as

$$\underline{\mathbf{R}}_0 = p(\omega_1)\,\underline{\mathbf{R}}_1 + p(\omega_2)\,\underline{\mathbf{R}}_2 \qquad (4.1.1)$$

where $p(\omega_1)$ is the *a priori* probability of occurrence of class 1, $\underline{\mathbf{R}}_1$ is the autocorrelation matrix of class 1 data, and similarly for $p(\omega_2)$ and $\underline{\mathbf{R}}_2$.

Since both $\underline{\mathbf{R}}_1$ and $\underline{\mathbf{R}}_2$ are real and symmetric matrices $\underline{\mathbf{R}}_0$ is as well. Thus a (similarity) transformation matrix $\underline{\mathbf{P}}$ may be applied to $\underline{\mathbf{R}}_0$, such that

$$\underline{\mathbf{P}}\,\underline{\mathbf{R}}_0\,\underline{\mathbf{P}}^t = \underline{\mathbf{I}} \qquad (4.1.2)$$

where $\underline{\mathbf{I}}$ is the identity matrix, and superscript t (^{t}) is the transpose operator, as before.

By substituting $\underline{\mathbf{R}}_1$ and $\underline{\mathbf{R}}_2$ into Equation 4.1.2, the autocorrelation matrices are transformed to

$$\underline{\mathbf{S}}_i = \underline{\mathbf{P}}\,p(\omega_i)\,\underline{\mathbf{R}}_i\,\underline{\mathbf{P}}^t \qquad (4.1.3)$$

for i=1,2 respectively.

Using the definition of $\underline{\mathbf{R}}$ from Equation 2.3.2.2, the corresponding transformation to the the data matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ is

$$\underline{\mathbf{S}}_i = \underline{\mathbf{P}}\, p(\omega_i) \frac{1}{T} \underline{\mathbf{X}}_i \underline{\mathbf{X}}_i^t \underline{\mathbf{P}}^t$$

$$= \left(\sqrt{\frac{p(\omega_i)}{T}}\, \underline{\mathbf{P}}\, \underline{\mathbf{X}}_i \right) \left(\sqrt{\frac{p(\omega_i)}{T}}\, \underline{\mathbf{P}}\, \underline{\mathbf{X}}_i \right)^t \qquad (4.1.4)$$

for i=1,2.

For convenience the transformed data matrices for class 1 and 2 data are denoted by $\underline{\mathbf{F}}^{(1)}$ or $\underline{\mathbf{F}}^{(2)}$, where $\underline{\mathbf{F}}^{(i)}$ is given by

$$\underline{\mathbf{F}}^{(i)} = \sqrt{\frac{p(\omega_i)}{T}}\, \underline{\mathbf{P}}\, \underline{\mathbf{X}}_i \qquad (4.1.5)$$

A property of this transformation shown by Fukunaga and Koontz is that the basis functions of $\underline{\mathbf{F}}^{(1)}$ and $\underline{\mathbf{F}}^{(2)}$ are equivalent. In order to demonstrate this first note that from Equation 4.1.2

$$\underline{\mathbf{S}}_1 + \underline{\mathbf{S}}_2 = \underline{\mathbf{I}} \qquad (4.1.6)$$

If the original pattern vectors are n-dimensional each of the n eigenvalues (λ) and vectors (ϕ) for both pattern classes is given by

$$\underline{\mathbf{S}}_1\, \phi_j^{(1)} = \lambda_j^{(1)} \phi_j^{(1)} \qquad (4.1.7a)$$

$$\underline{\mathbf{S}}_2\, \phi_j^{(2)} = \lambda_j^{(2)} \phi_j^{(2)} \qquad (4.1.7b)$$

for j=1,2...n.

Assume that the eigenvalues and corresponding eigenvalues for class 1 data are ordered in an descending fashion, as per Equation 2.3.3.24.

Using Equation 4.1.6 the eigenvalues and vectors for class 2 data can also be expressed as

$$\underline{\mathbf{S}}_2\, \phi_j^{(2)} = (\underline{\mathbf{I}} - \underline{\mathbf{S}}_1)\, \phi_j^{(2)} = \lambda_j^{(2)} \phi_j^{(2)} \qquad (4.1.8)$$

which can be written after multiplication and regrouping as

$$\underline{\mathbf{S}}_1\, \phi_j^{(2)} = (1 - \lambda_j^{(2)})\, \phi_j^{(2)} \qquad (4.1.9)$$

Comparing this result with Equation 4.1.7a the following two important relationships emerge

$$\phi_j^{(1)} = \phi_j^{(2)} \tag{4.1.10a}$$

$$\lambda_j^{(1)} = 1 - \lambda_j^{(2)} \tag{4.1.10b}$$

The first result demonstrates that the eigenvectors from both pattern classes are equivalent, the second that the eigenvalues for each pattern class are reversely ordered and lie between 0 and 1.

It can easily be shown that $\lambda_j^{(1)}$ and $\lambda_j^{(2)}$ are both less than or equal to unity and are reversely ordered. First note that the $\lambda_j^{(i)}$'s are the eigenvalues of $\underline{\mathbf{S}}_i$, and therefore represent the power in the rows of $\underline{\mathbf{F}}^{(i)}$. Since the $\underline{\mathbf{F}}^{(i)}$'s are real the eigenvalues will be real and positive. Note also that Equation 4.1.6 stipulates that the sum of the jth diagonal elements from $\underline{\mathbf{S}}_1$ and $\underline{\mathbf{S}}_2$ is equal to 1. In order to satisfy the fact that the eigenvalues are positive and sum together to 1, the $\lambda_j^{(i)}$'s must be less than or equal to 1. Combining this result with the descending order imposed on the the eigenvalues for class 1 it is easy to see that the eigenvectors which account for the largest variance in class 1 account for the smallest in class 2, and visa versa.

This last point is useful for classification. For example, if a transformation matrix was created from a training data set, the first step to classify new data would be to use Equation 4.1.5 and project the result, i.e. $\underline{\mathbf{F}}^{(i)}$, on to the space defined by the training set basis vectors. By noting which basis functions accounted for the largest variance a decision could be made as to which category the unknown data belonged. Ideally, the training set would have as much variance of the data compressed into as few basis functions as possible, and thus only the highest and lowest would be required for classification. Otherwise, all basis functions would be required and some form of distance metric needed to measure the similarity of power distribution in the training set data to the unknown samples.

In this thesis all basis functions were used in conjunction with the minimum distance classifier (Gevins, 1987), based on Euclidean distances. After the FKT was applied to a training set consisting of data from both classes (to create $\underline{\mathbf{R}}_0$), two vectors containing the 31 eigenvalues for each pattern class was created. Since it was known the corresponding eigenvectors were orthogonal, each vector therefore defined a point in 31 dimensional space. The classification proceeded by transforming the "unknown" data using Equation 4.1.5, and then projecting the result on to the training set eigenvectors (basis functions) creating a 31 dimensional vector; another point in the measurement space. The Euclidean distance between this point and the two training set vectors (points) was then computed. The class of the unknown sample was assigned to same category as the training set vector yielding the smallest distance.

An interesting property arouse from using all of the basis vectors from the transformed space. Since the Euclidean distance between two points was constant, regardless of the set of basis functions used, the minimum distance classifier described above gave the same results for any set of basis functions defining a 31 dimensional space. Viewed another way: the basis functions defined by the FKT were the set accounting for the largest amount of power, for both classes, in the transformed space. Since the minimum distance approach did not utilize this information the exact basis set was unimportant. What was effectively measured using all of the basis functions, therefore, was how well class data was separated in the measurement space by the transformation matrix.

To verify that the transformation matrix adequately separated class data three routes of investigation were taken. The first, and simplest, described in Section 4.2, used a sample of data from each class from an individual to create a transformation matrix. This matrix was then applied to "unknown" data from that subject to determine how well these new vectors clustered around the training vectors used in creating the transformation matrix. The next test, described in Section 4.4, used a transformation matrix from one subject to determine class separation, and the classification rate, of data from other subjects. The final test performed, described in Section 4.6, used data from several subjects to create the transformation matrix, with class separation and classification rate measured on data from those as well as other subjects. In all of these tests the programs used to compute the transformation matrix, and perform classification assumed that the *a priori* probabilities for each class were 0.5.

The transformation matrix $\underline{\mathbf{P}}$ was solved using the following relation

$$\underline{\mathbf{P}} = \sqrt{\frac{1}{\underline{\lambda}}}\,\underline{\mathbf{U}}^t \tag{4.1.11}$$

where $\underline{\mathbf{U}}$ is the eigenvector matrix of $\underline{\mathbf{R}}_0$, (the class combined autocorrelation matrix defined in Equation 4.1.1), and $\sqrt{1/\underline{\lambda}}$ is a diagonal matrix containing the reciprocals of the square roots of the eigenvalues of $\underline{\mathbf{R}}_0$.

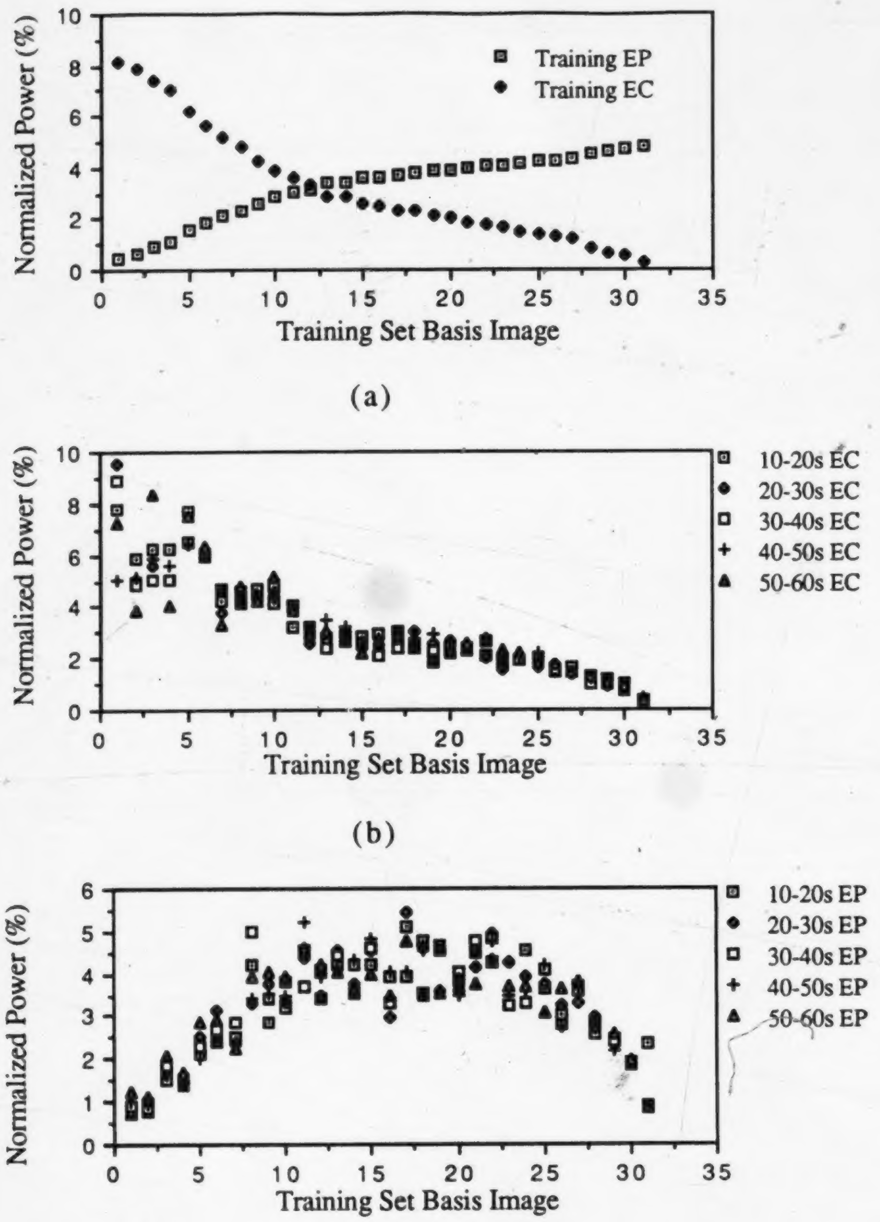
4.2 Classification on Individual Data - Methods and Results

In this section the results of classification are presented for individual data. For each of the subjects used in Chapter 4 training set data was taken from the first 10 seconds of the subject's EC and EP records. To avoid effects of d.c. amplifier bias and problems associated with signal power between states, each electrode sequence was made zero mean, and each image normalized to unit power before FKT analysis. The transformation matrix, **P**, training set basis functions (eigenvectors) and training vectors (eigenvalues) were then computed and stored.

To test the classification ability five 10 second segments of data from each pattern class record were used, with each segment first pre-processed as above. The transform matrix was then applied via Equation 4.1.5 to create a set of transformed data matrices. Each data matrix was projected on to the training set basis functions (using Equation 3.4.5) which generated a set of coefficients for each basis function. Next, the normalized power represented by each training set basis function was measured (using Equation 3.4.1), with all 31 such values taken together to form a pattern vector, that is, a point in 31 space. The set of pattern vectors corresponding to the transformed data matrices were then each classified using the minimum distance classifier, as described above.

Figure 4.2.1 displays the results for Subject 1. Graph (a) shows the training set pattern vectors, while graphs (b) and (c) exhibit the feature vectors when transformed EC and EP data are projected on to the training set basis functions. Again, normalized power values are used to measure the results of these projections. Table 4.2.1 shows the Euclidean distance between the feature vector and training set vectors, the classifiers output, and true class of data.

For each of the other four subjects used in Chapter 3 the same classification procedure as above was performed, with the results summarized in Table 4.2.2.



(a)

(b)

(c)

Figure 4.2.1

Pattern Classification for Subject 1 Using Training Data from Subject 1: (a) Training Set Feature Vectors (b) Feature Vectors from Five 10 Second Segments of EC Data (c) Feature Vectors from Five 10 Second Segments of EP Data

Data Section	Distance to EC Training Vector	Distance to EP Training Vector	Class of Data	Classifier Output
10-20s	3.19	17.40	EC	EC
20-30s	4.99	17.79	EC	EC
30-40s	5.16	17.30	EC	EC
40-50s	5.26	15.55	EC	EC
50-60s	6.17	17.18	EC	EC
	Mean = 5.0	Mean = 17.0	Classification Rate =	
	Std. = 1.1	Std. = 0.9	5/5 = 100%	
10-20s	17.11	6.23	EP	EP
20-30s	16.58	7.02	EP	EP
30-40s	16.29	7.28	EP	EP
40-50s	16.71	7.15	EP	EP
50-60s	15.53	7.10	EP	EP
	Mean = 16.4	Mean = 7.0	Classification Rate =	
	Std. = 0.6	Std. = 0.4	5/5 = 100%	

Table 4.2.1
Classification Results For Subject 1

Subject	Data Class	Classification Rate	Mean Distance, Standard Dev. to EC Training Vector	Mean Distance, Standard Dev. to EP Training Vector
2	EC	80 %	7.2, 1.2	9.6, 1.1
	EP	100 %	10.4, 0.5	5.0, 0.4
3	EC	100 %	4.3, 0.6	10.4, 0.7
	EP	100 %	8.9, 0.9	5.9, 1.1
4	EC	100 %	5.8, 0.8	13.8, 0.9
	EP	100 %	16.9, 1.0	4.1, 0.8
5	EC	100 %	4.3, 0.2	13.0, 0.5
	EP	100 %	10.4, 0.3	7.7, 0.1

Table 4.2.2
Classification Results For Subjects 2 - 5

4.3 Classification on Individual Data - Discussion

The results in Section 4.2 demonstrate that for two class data from an individual recorded during the same session, accurate supervised classification between cognitive states can be performed using the Fukunaga-Koontz transformation. Moreover, this accuracy is obtained with a simple classifier, using only single training samples of data from each class.

The training vectors (amount of power represented by each basis function after FKT normalization of the training data) are shown in Figure 4.2.1a. The main point to note is that for both classes the power represented in the basis space is not concentrated in a few basis functions, as was the case for non-normalized data. The shape of the graphs (location in 31 space), although different, resemble each other enough to indicate that they both are located in similar subspaces of the measurement space. In other words, the FKT is not separating the two classes into two distinct regions of 31 space: in the ideal case the first and last basis functions would account for all of the power in the first and second data classes, respectively. The non-ideal separation of the two training vectors is further evidence that large similarities exist between the data classes, a result consistent with the findings from Sections 3.4.1 and 3.5.1.

The projections of normalized data on the basis training space, shown for EC and EP data in Figures 4.2.1(b) and (c) respectively, demonstrate that the feature vectors are located close to the training set vectors, with the EP data exhibiting more fluctuation than the EC data. This may indicate more spatial variation of the EEG exists within the EP state than the EC state, or that artifacts, more common in EP data, are present. While variation does exist in both classes, the general trend is that the transformation matrix separates pattern class data from other portions of the record, as far as one minute after the training vector data. This is consistent with the results in Chapter 3, where little variation in the fundamental spatial patterns of the EEG was found over long periods.

The classification results for Subject 1, shown in Table 4.2.1, demonstrate correct classification of the unknown data for all cases tested. This again indicates that the basic spatial patterns within a cognitive state are maintained for periods of at least one minute, and that the normalization matrix separates the data class's feature vectors far enough apart for adequate discrimination. Although not shown in the table, the inter-class distance between the training vectors is 19.4. This value combined with the average distance measure between the feature and training vectors, and the standard deviations, show that the classes are adequately separated, and form compact "clusters" in 31 space.

The latter point is important, indicating that more sophisticated classifiers based on cluster analysis techniques can be used. For example, rather than having only a single training vector, all previous correctly classified feature vectors can used, with distance to the either the nearest neighbour of each cluster, or mean vector from each cluster, the basis for classification. Needless to say, these methods are more computationally sophisticated, but also provide better performance.

For the other subjects classification accuracy is also excellent: 100% for all but one subject. The ratio of inter to intra-class distance for most subjects is again high, indicating good class separation. The inter-class distance of Subject 2, for whom misclassification occurred, is smaller than that of the other subjects, indicating that both pattern classes in this subject are very similar. The strip chart recordings from Subject 2 confirm that little difference between the eyes closed and open recordings exist.

Several important results are shown in this section. The first is that for intra subject data representing the EC and EP cognitive states the FKT is useful in transforming the measurement space so that the resulting feature vectors can be used for classification. These vectors cluster in the transformed space such that the inter-class distance is larger than the intra-class distance, a useful property for using more sophisticated classifiers. Finally, when class differences are small, as determined by visual inspection of the strip chart recordings, the inter-class distances decrease resulting in a lower classification rate.

4.4 Classification Across Individuals - Methods and Results

The last section demonstrated that classification between cognitive states can be performed on intra subject data: the next question asks if classification can be performed on inter subject data. That is, can the transformation matrix from one individual be used to classify data from another subject?

In asking this two approaches are possible. In the first method training vectors are taken from the subject whose transformation matrix is used, while in the second, training samples from the subject whose data is to be classified are used. If good classification occurs using the former approach then the pattern vectors from the new data have aligned themselves with those of the first subject. This indicates that not only are the features extracted by the FKT to classify between states for the two subjects similar, but also that the vectors cluster in the same locations within the measurement space. If this is the case for several subjects then the FKT measurement space and transformation matrix may provide a universal method of classification, otherwise individual differences will have to taken into account, via the latter approach.

Table 4.4.1a shows the results of classification when the transformation matrix and training vectors from Subject 1 were used to classify 5 ten second segments of data, starting from 10 seconds into the record, from each of the other 4 subjects. Table 4.4.1b shows similar results, except that training vectors taken from the first ten second segment of the EC and EP records from the subject to be classified were used. Included in the tables is the average distance to each training vector, and the standard deviation of those distances, over the 5 classification attempts.

Subject	Data Class	Classification Rate	Mean Distance, Standard Dev. to EC Training Vector	Mean Distance, Standard Dev. to EP Training Vector
2	EC	0 %	14.1, 1.1	11.8, 0.8
	EP	100 %	16.5, 1.6	9.8, 0.6
3	EC	0 %	20.5, 0.5	9.3, 0.6
	EP	100 %	17.2, 0.6	8.5, 0.9
4	EC	0 %	14.7, 0.3	12.5, 0.3
	EP	100 %	17.2, 0.6	8.5, 0.9
5	EC	0 %	13.1, 0.7	12.5, 0.8
	EP	100 %	16.9, 0.8	9.4, 0.6

(a)

Subject	Data Class	Classification Rate	Mean Distance, Standard Dev. to EC Training Vector	Mean Distance, Standard Dev. to EP Training Vector
2	EC	80 %	4.9, 1.1	5.7, 1.0
	EP	100 %	6.9, 1.4	3.7, 1.1
3	EC	100 %	2.9, 0.6	6.1, 0.5
	EP	100 %	5.3, 1.2	3.8, 1.0
4	EC	100 %	3.4, 0.3	8.7, 0.4
	EP	100 %	11.0, 1.0	3.8, 1.0
5	EC	100 %	3.8, 0.9	8.7, 0.6
	EP	100 %	5.6, 0.5	4.2, 0.5

(b)

Table 4.4.1

Classification Results Using Measurement Space and Transformation Matrix of Subject 1: (a) Training Vectors From Subject 1 (b) Training Vectors From Subject Being Classified

4.5 Classification Across Individuals - Discussion

The small standard deviation values compared with the large mean distance values shown in Table 4.4.1a indicate that, as before, each subject's transformed data is clustering. The large mean values, compared with those computed in Section 4.2, listed in Table 4.2.2, show that the clusters are distributed further away from the training vectors from Subject 1, than from their own training vectors. Furthermore, the mean values show that the clusters are all located closer to Subject 1's EP training vector than to the EC training vector, resulting in the 100% misclassification rate for the EC data.

On the other hand, when training vectors from the subject being studied are used, as shown in Table 4.4.1b, the classification rate is high and exactly matches that from Section 4.2, where the subject's own transformation matrix was used. Although this is the case, better class separation occurs in the latter as shown by comparing the mean distances to the training vectors listed in Table 4.4.1b and Tables 4.2.1 and 4.2.2. In other words, the differences between each class vary on a per subject basis, a finding consistent with the results of Chapter 3.

It is possible that a source for the low classification rates found in Table 4.4.1a is the choice of the training vectors (i.e. always taking the first 10 second segment of Subject 1's data), and all that is required to raise the classification rate is the use of other segments of the data. While this has not been attempted, the compactness of the clusters from each state indicate that while the precise position of the training vector may change in using other data segments, the shift would not be great enough to alter the classification rate. It is therefore for convenience that the first ten second segment of each subject's data record is used to compute the training vector throughout this Chapter.

To summarize then, these results demonstrate that individual differences are important in classification between cognitive states. This manifests itself in the fact that, when using another subject's transformation matrix, examples of class data from the individual being studied are required to classify other data from that subject. With such training samples the classification rate matches that when the transformation matrix computed from each subject is used, although class separation is not as high as in the latter case. In both cases, the transformed data forms clusters in the measurement space.

4.6 Classification Using Collective Data - Methods and Results

The final route of investigation focuses on using the data from several subjects to create the transformation matrix. The goal in such an operation is to smooth out individual differences and create a matrix to separate EC/EP class data for most individuals, without a need for training vectors from each subject to be classified.

The first 10 second segment of data from each of the five subject's EC and EP record was used to create the collective data set. As in Section 3.5, each subsection of data was made zero mean temporally, and then each image made unit power, in order to avoid the effects of individual differences in EEG power. After computation of the transformation matrix five 10 second segments of data from each subject for each class were classified. A set of ten training vectors were used for classification, each corresponding to a data subsection from the collective data set.

With more than two training vectors several methods based on the minimum distance classifier are possible. The first computes the distance to all of the 10 training vectors, with the classifier output being the class of training vector with the smallest distance. The second approach reduces the number of training vectors from 10 to 2, by averaging the vectors from each class, and then applies the minimum distance classifier to the two training vectors.

The results obtained using both of the above classification methods are listed in Table 4.6.1.

Subject	Data Class	Classification Rate	Mean Distance, Std. to Min. Distance EC Training Vector	Mean Distance, Std.to Min. Distance EP Training Vector
1	EC	100%	4.7, 0.8	12.2, 1.0
	EP	100%	10.4, 0.5	6.3, 0.6
2	EC	80 %	4.6, 1.6	7.6, 2.0
	EP	100 %	7.6, 1.0	2.9, 0.9
3	EC	100 %	3.5, 0.8	5.2, 0.6
	EP	100 %	6.5, 1.0	3.6, 1.2
4	EC	100 %	4.7, 0.9	7.9, 0.6
	EP	100 %	11.7, 0.8	2.9, 0.6
5	EC	100 %	3.5, 1.0	9.3, 0.6
	EP	100 %	7.8, 0.3	4.4, 0.5

(a)

Subject	Data Class	Classification Rate	Mean Distance, Standard Dev. to Mean EC Training Vector	Mean Distance, Standard Dev. to Mean EP Training Vector
1	EC	100%	6.5, 0.9	15.3, 1.0
	EP	100%	11.3, 0.9	5.3, 0.6
2	EC	100 %	6.2, 0.2	10.9, 2.1
	EP	60 %	7.4, 0.8	6.4, 0.9
3	EC	0 %	10.5, 0.4	7.8, 0.9
	EP	100 %	9.6, 0.7	5.4, 0.5
4	EC	100 %	6.4, 1.0	10.2, 0.5
	EP	100 %	13.6, 0.7	5.2, 0.9
5	EC	100 %	4.7, 0.8	12.0, 0.6
	EP	100 %	9.2, 0.4	4.7, 0.4

(b)

Table 4.6.1

Classification Results Using Transform Matrix Based On Collective Data Set : (a) Minimum Distance Classifier Using 10 Training Vectors (b) Minimum Distance Using Mean Training Vectors

A problem when working with many variables is that it is often difficult to picture the position and shape of the pattern vectors in the measurement space. A common technique is to project the data on to a two dimensional space (plane), where as much information as possible about the orientation of the vectors is maintained. Many methods have been proposed for such purposes (Biswas et al., 1981), some transformations being linear, others non-linear. In order to provide further insight into the reasons for both classification and misclassification of the EEG pattern vectors, a projection algorithm was used here.

In the keeping with the spirit of this thesis, the Karhunen-Loeve transform was chosen. First, the entire set of pattern vectors (both training and feature vectors) were taken together to form a data matrix, and the KLT applied. The normalized dot product (projection) of each pattern vector on to the first and second basis functions (eigenvectors) was then computed. A scatter plot was finally created for all pattern vectors, with the abscissa the projection on basis function 1, and the ordinate the projection on basis function 2.

This method was applied to the set of 60 pattern vectors (6 per subject, per state) generated above. The first and second principal components accounted for 85.9%, and 7.7% of the power in the original data, respectively. Figure 4.6.1 presents five scatter plots, one for each subject, showing the clustering effects of each subject's EC and EP data. Note that all of these plots are at the same scale.

In order to more clearly demonstrate clustering for all five subjects Figure 4.6.2 shows a composite of the above plots.

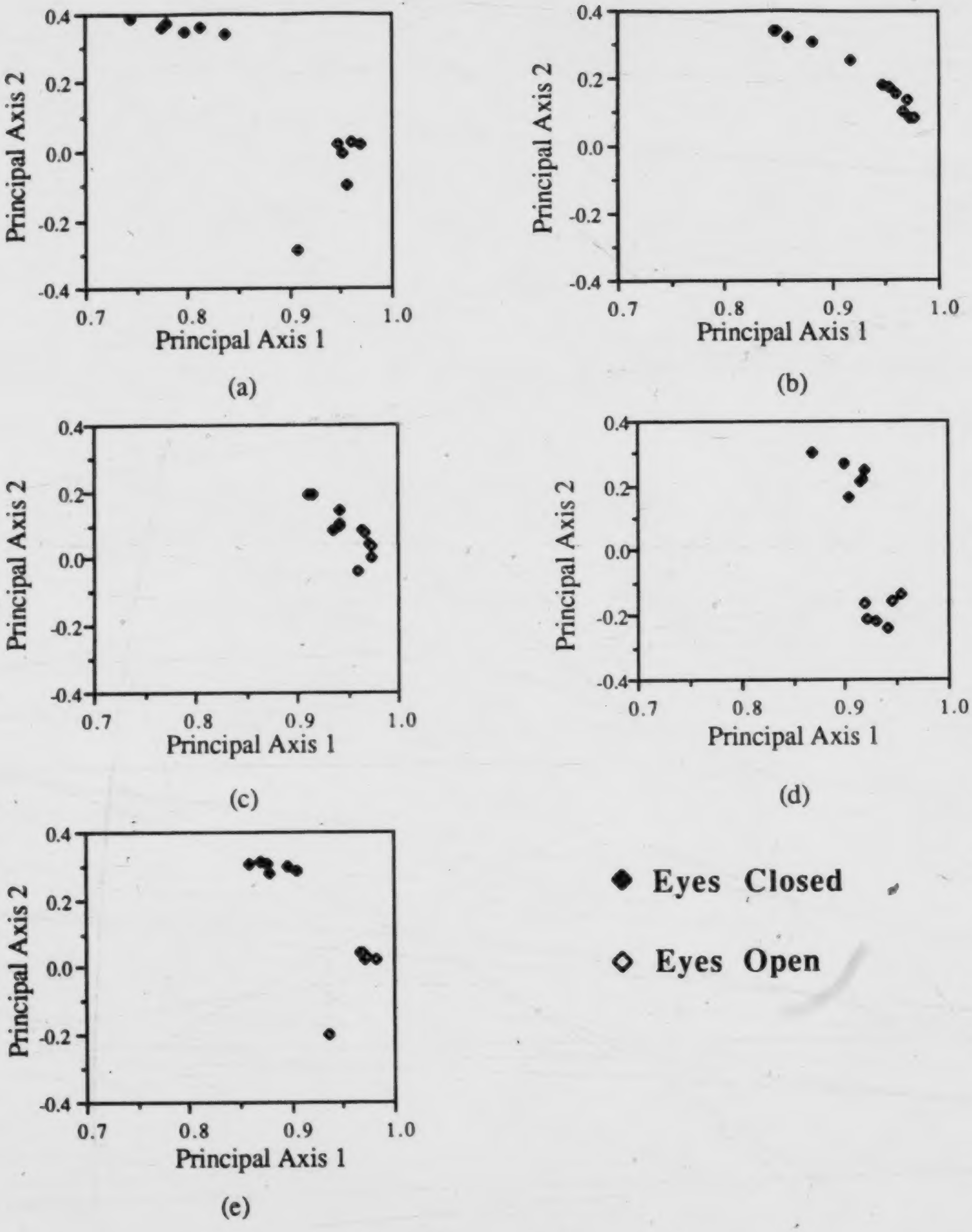


Figure 4.6.1

Normalized Projection of EC and EP Pattern Vectors on to the First Two Principal Component Axes of the Feature Vector Set (a) Subject 1 (b) Subject 2 (c) Subject 3 (d) Subject 4 (e) Subject 5

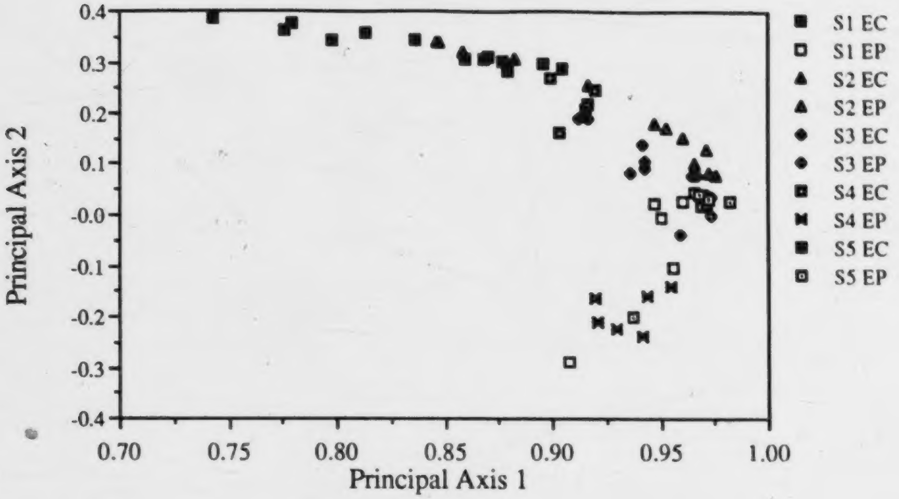
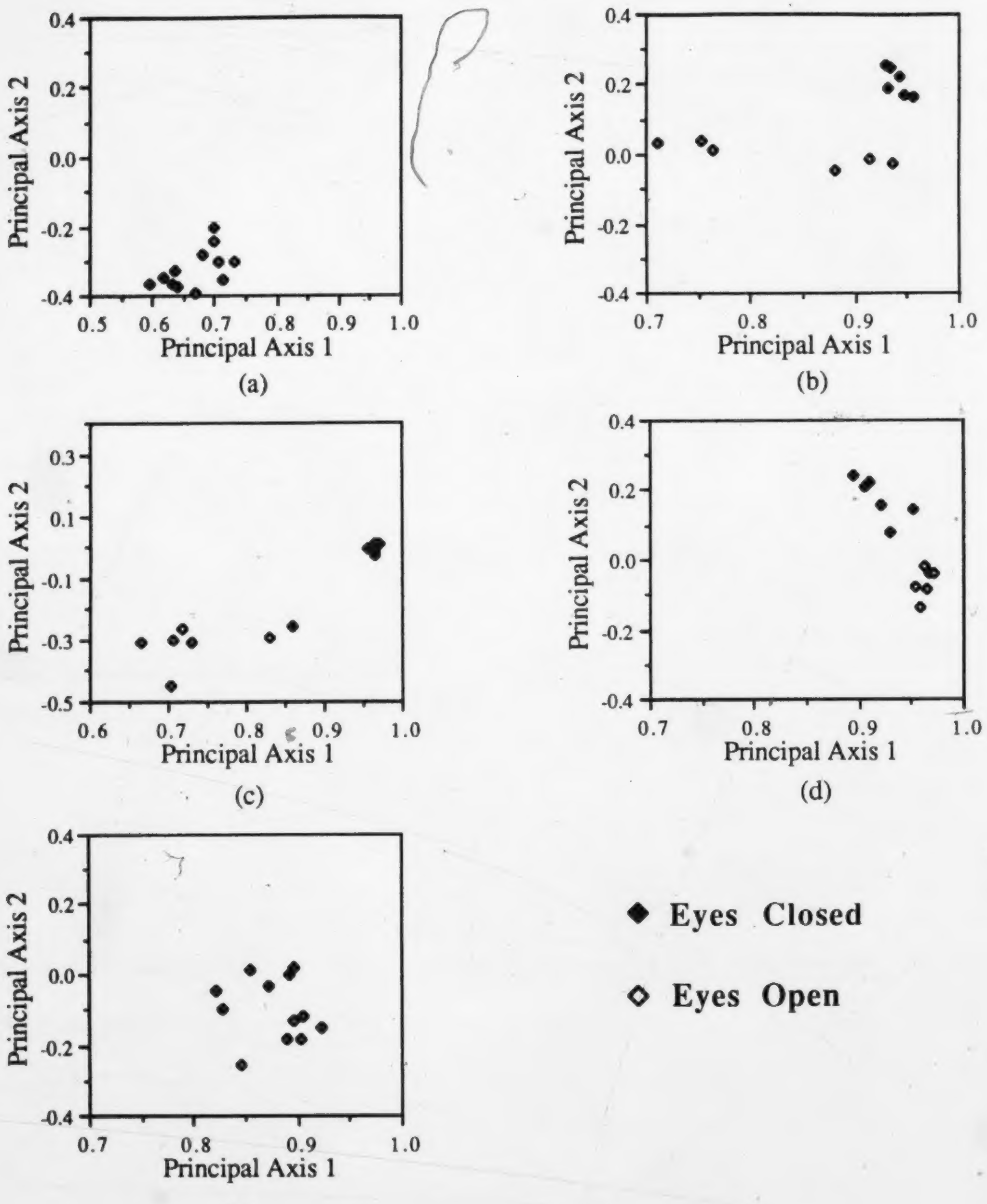


Figure 4.6.2

Composite Projection of EC and EP Pattern Vectors on to the First Two Principal Component Axes of the Feature Vector Set

The purpose of using a transformation matrix based on collective data is to see if other individuals data can be classified using the training vectors from the subjects in the collective data set. In other words, do the pattern vectors from other subjects cluster in similar locations as those of the subjects from the collective set? An easy way to view such data is with the two dimensional projection method described above.

Figure 4.6.3 shows the results when data from an additional 5 subjects, denoted Subject 6 through Subject 10, was transformed by the collective transformation matrix, and then projected on to the two principal component axes used in Figures 4.6.1 an 4.6.2. Table 4.6.2 gives the classification rate for each of these subjects using the first ten seconds of their EC and EP record for the training vector data.



(a) (b) (c) (d) (e)

Figure 4.6.3

Normalized Projection of Transformed Data on to Principal Component Axes Used in Figures 4.6.1 and 4.6.2
(a) Subject 6 (b) Subject 7 (c) Subject 8 (d) Subject 9
(e) Subject 10

Subject	Data Class	Classification Rate
6	EC EP	100% 80%
7	EC EP	100 % 100 %
8	EC EP	100 % 100 %
9	EC EP	100 % 100 %
10	EC EP	100 % 100 %

Table 4.6.2
Classification Results Using Subjects Own Training Vectors and Collective Transformation Matrix - Subjects 6-10

4.7 Classification Using Collective Data - Discussion

As indicated by the results from Section 4.4, classification between states using the FKT and a minimum distance classifier requires training data from each subject to be classified. In other words, the methods used here do not find a universal transformation matrix and set of training vectors to differentiate between the EC and EP states, but rather ones which are useful only for the subjects whose data was used to create the training data.

The collective transformation matrix clusters each subjects' data, as shown in Figures 4.6.1 and 4.6.2. For some subjects the inter to intra-class distance ratio is higher than for others, indicating again that some subjects exhibit significant similarity in their EC and EP data, while others do not. For example, Subjects, 1, 4 and 5 demonstrate large inter-compared to intra-class distance, while for Subjects 2 and 3 the EC and EP clusters are much closer together.

This variation amongst the individuals leads to the problem of selecting the training vectors. Selecting only one example from each class might easily result in misclassification, particularly since there is a region where both EC and EP vectors are close together. Selecting the data used in creating the transformation matrix, as was done in Table 4.6.1a, is a good approach, but in this case requires that 10 distance computations be performed. Furthermore, the minimum distance training vector invariably turns out to be from the same subject; this is reflected by the fact that the same classification accuracy obtained was exactly that using individual data (Section 4.2).

Using averaged training vectors simplifies the computational load on the classifier, but results in reduced accuracy for subjects whose data is close to the halfway point of the mean vectors, as shown in Table 4.6.1b. For Subject 3, whose classification rate was high using all 10 training vectors, the accuracy fell to 0% for the EC data, as each of the EC vectors was closer to the EP mean vector. Subject 2's EC data was classified at 100%, whereas the EP rate fell to 60%, demonstrating that Subject 2's EC and EP data resembled the typical EC data from the entire subject group. Both of these results again indicate that training vectors from the subject being classified provide the highest classification accuracy.

Although not perfect, the collective transformation matrix, **P**, adequately separates the EC and EP data and results in good classification accuracy from the individuals whose data was used in creating **P**. The main question posed in this section, though, is how well does **P** separate data from other individuals. The two dimensional projections of five other subjects' data, shown in Figure 4.6.3, demonstrate that while the transform matrix may, in some cases, separate the classes, the location of the clusters is not similar to the original group of subjects. This again confirms that differences between the EC and EP states are dependent upon the individual.

For example, the feature vectors from Subject 6 all cluster closely together, and in positions where none of the vectors from Subjects 1-5 are located. The data from Subjects 7-10, while clustering by cognitive state, show no trend in the location of the clusters, and again are not aligned with those from Subjects 1-5. It is obvious therefore, that to provide any sort of classification whatsoever, training vectors from the subject to be classified are required. Furthermore, Figure 4.6.3 demonstrates that, if classification were based only on the two axes shown, a judicious choice of training vectors would be required for each subject as the intra-class distances for each state are comparable to the inter-class distances.

A point worth noting is that the axes that Subjects 6-10 feature vectors have been projected on to is optimal only for the training vectors from Subjects 1-5. Thus, the scatter plots may not fully represent the inter and intra-class distances of the feature vectors. The purpose of the figure is to show if the new pattern vectors cluster in the same locations as those from Subjects 1-5, which they obviously do not. Classification though, is based on using the entire 31 dimensions and is thus not prone to the above distortion.

Table 4.6.2 shows the classification results when, as before, the feature vectors from the first 10 second segment of the EC and EP records are used as training vectors. Although not perfect, the results do show these vectors provide reasonable classification accuracy using the collective transformation matrix. This also implies that the scatter plots in Figure 4.6.3 do not adequately capture the location of

Subject's 6-10 pattern vectors.

The results from this section confirm those from Section 4.4, which state that classification using the FKT requires training samples from the subject being classified, regardless of the transformation matrix used. Moreover, the results are in agreement with those from Chapter 3, where differences between states were detectable using the KLT coefficients, but where no general trends were found amongst the individuals studied. As in Chapter 3, finer division of the subject groups may help to illuminate any such trends. That is, do the feature vectors from subjects of the same sex, age, etc., cluster at similar locations in the transformed measurement space?

4.8 Discussion and Conclusions

The main result from this chapter is that classification between cognitive states within an individual can be performed with the Fukunaga-Koontz transformation and a simple minimum distance classifier. The need for further investigation is clearly pointed out by the lack of accuracy when applying the classification parameters from one subject (or even a group of subjects) to those of others.

The results from Chapter 3 demonstrated that differences existed in the proportion of basis functions present in the EC and EP states within a subject. The FKT provides a method to quantify these differences in the form of feature vectors, which may be input to a simple classifier. Chapter 3 also indicated that while differences between states existed, no general trend was found in the subjects studied. This fact is reflected in the poor classification accuracy when the classification parameters from one subject are applied to other subjects. It is difficult therefore, to ascertain whether the transformation matrix is quantifying physiologically related differences, or differences related to the recording session or other such factors.

Due to this latter point it is difficult to speculate if the FKT would be useful for detection of brain dysfunctions which manifest themselves in the EEG. For example, for a patient with a tumour it would be difficult to detect the abnormality with the FKT unless a "non-tumour" example of that patient's EEG were available. For epileptics, on the other hand, non-epileptic EEG is available and therefore detection and possibly warning would be possible. Because the classification performed here is only on two states, EC and EP, research is required to determine the applicability of the FKT to other classes of the EEG data.

Further investigation is also required to determine if the feature vectors have any physiological basis, or if they are just convenient mathematical entities lacking any intuitive appeal. Using simulated data, or data collected from well defined subject groups, and noting where the transformation matrix clusters the classes would be useful for understanding the transformed measurement space. For example, data generated from a front/back dipole may cluster in a different location than that of data from a side/side dipole.

In general then, the results from this chapter point in one of two directions. The first, is that while differences between cognitive states are detectable within an individual, no general trend exists amongst the population. The second, is that tighter experimental controls are required to determine if any trends can be found using the FKT.

Chapter 5

Conclusions

In this thesis the Karhunen-Loeve transform was used as a pattern representation and classification technique on the spatio-temporal EEG. The study used background EEG data and concentrated on showing that the results obtained were universal in nature, in so far as the methods yielded similar findings on every subject studied. The parameters themselves, though, were unique to every individual.

When applied to a set of data the KLT yields a set of basis functions and coefficients. These entities have two desirable properties. First, maximum information is compressed into the fewest number of basis functions and coefficients possible, in a least squares sense. Second, the set of basis functions and coefficients both form an orthogonal set. When applied to the spatial EEG the basis functions represent the spatial patterns underlying the original input data, and can thus be considered basis images. In such a case the coefficients represent the amount of each basis image present in the original data at various points in time.

A key result when the KLT was applied to temporal image sequences of EC and EP data was that the original data could be reconstructed typically using 5 basis images. Moreover, the basis images computed from one section of data adequately represented other segments of data separated by as much as 90 seconds. The representation was accurate both in terms of power and shape. These results indicate that significant compression of EEG data with little loss of signal power and shape is possible using the KLT. This is consistent with the results from other studies (Nunez, 1981; Skrandies and Lehmann, 1982), where physiological relevance of the basis images was also postulated.

The amount of spatial variation, reflected in the quantity of each basis image present, varied with the length of data analyzed. Short data segments (less than 1/2 second) exhibited less spatial variation than longer segments. The maximum amount of variation occurred at data lengths of approximately 8 seconds, after which little increase occurred.

The basis images and the amount of power represented by each image varied between individuals and cognitive states. For the states studied, front back activity was usually found in the first two basis images, with the third typically containing side to side activity. The ratio of power represented by each basis image varied through each subject's data record, and was very different among the subjects themselves. There was more commonality, though, in the basis images of subjects within the same cognitive state, than for basis images from two different states of the same subject.

In order to quantify the differences between EC and EP data the Fukunaga and Koontz transformation and a minimum distance classifier was used. The FKT requires as input two training samples of data, one from each class. The output consists of a transformation matrix, which when applied to unknown data yields feature vectors, and when applied to the training data yields two training vectors. The classifier assigns unknown feature vectors to the same class as the training vector to which it is closest (using a Euclidean distance).

The FKT accurately classified unknown data when training vectors from the subject to be classified were available, regardless of the subject whose data was used in computing the transformation matrix. Better class separation, though, occurred when the transformation matrix and training vectors were from the same subject. Differences between cognitive state were therefore dependent upon the individual, and were not common amongst the subjects studied. Using collective data to generate the transformation matrix did not improve the classification accuracy for other subjects (i.e. those whose data was not used in computing the matrix), unless training vectors from those other subjects were available. These results are consistent with the fact that the amount of each basis image present in the EEG varied considerably for subjects in the same state.

The lack of any uniform trend in the pattern representation or classification results may be indicative of several factors. First, many more subjects were required in order to smooth out statistical fluctuations to detect any significant trends, or that finer division along subjects groups was required, or both.

Second, while the cortical processes generating the EEG are similar for subjects in the same state, other factors such as scalp conductivity and skull thickness were distorting the signals so that they appeared very different on the scalp. Modelling such distortion may uncover the "true" signals, which then may be input to KLT with more success.

Finally, individual variations in the processes generating the EEG were sufficiently different so that no common framework exists. This is the view that Gevins (1984) supports when he states, "... wide variations in fundamental characteristics of the time series (frequency, amplitude, morphology, spatial distribution, etc.) due to normal anatomical, and biophysical differences have prevented the development of precise quantification of most wave properties" (p. 834). If this is the case, then until truly normative values can be obtained from the EEG, techniques such as the KLT will have to be applied only to intra-subject data.

Generally, the results from this thesis provide a basis for further investigation into the EEG using the Karhunen-Loeve transform. While the KLT will by no means explain all remaining questions regarding the EEG, it is hoped that the methods developed here are used to further work in the field by providing an additional analytic tool.

REFERENCES

H. Berger, "Uber das Elektrenkephalogramm des Menschen," *Arch, Psychiat. Nervenkrankheiten*, vol. 87, pp. 527-570, 1929.

G. Biswas, A.K. Jain, and R.C. Dubes, "Evaluation of Projection Algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-3, No. 6, pp. 701-708, 1981.

R.M. Chapman, J.W. McCrary, H.R. Bragdon, J.A. Chapman, "Latent Components of Event-Related Potentials Functionally Related to Information Processing," in *Progress in Clinical Neurophysiology*, J.E. Desmedt, Ed., Vol. 6, Cognitive Components in Cerebral Event-Related Potentials and Selective Attention, Basel, Switzerland: Karger, 1979, pp. 80-105.

B.E. Cooper, *Statistics for Experimentalists*, Braunschwieg, West Germany: Pergamon,1969.

E. Donchin and E.F. Heffley, "Multivariate Analysis of Event-Related Potential Data: A Tutorial Review," in *Proc. 4th Int. Congr. Event-Related Brain Potentials. Res. (EPIC IV)*, 1978, pp. 555-572.

R.A. Fischer, "On the Probable Error of a Coefficient of Correlation From a Small Sample," *Metron*, Vol. 1,1921.

K. Fukunaga, *Introdution to Statistical Pattern Recognition.* New York: Academic, 1972.

K. Fukunaga and W.L.G. Koontz, "Application of the Karhunen-Loeve Transform to Feature Selection and Ordering," *IEEE Trans. Comput.*, vol. C-19, pp. 311-318, 1970.

K. Fukunaga, "Statistical Pattern Classification," in *Handbook of Pattern Recognition and Image Processing*, T. Y. Young and K.S. Fu, Eds. Orlando, Florida: Academic, 1986, pp. 3-32.

A.S. Gevins, "Analysis of the Electromagnetic Signals of the Human Brain: Milestones, Obstacles, and Goals," *IEEE Trans. Biomed. Eng.*, vol. BME-31, No. 12, pp. 833-850, 1984.

A.S. Gevins, "Chapter 17: Statistical Pattern Recognition," in *Methods of Analysis of Brain Electrical and Magnetic Signals, EEG Handbook* (revised series. Vol. 1), A.S. Gevins and A. Remond, Eds. New York:Elsevier/North-Holland, 1987, pp. 541-582.

I.N. Gibra, *Probability and Statistical Inference for Scientists and Engineers*, Englewood Cliffs, N.J: Prentice Hall,1973.

E. Glaser and D. Ruchkin, *Principles of Neurobiologic Signal Analysis*. New York: Academic, 1976.

K. Hanakata, "Feature Selection and Extraction For Decision Theoretic Approach and Structural Approach," in *Pattern Recognition Theory and Application*, K.S. Fu and A.B. Whinston, Eds. Leyden, Holland:Noordhoft, 1977, pp. 133-169.

R.D. Katznelson, "Normal Modes of the Brain: Neuroanatomical Basis and a Physiological Theoretical Model," in *Electric Fields in the Brain: The Neurophysics of EEG.*, P. Nunez, New York:Academic, 1981, pp. 400-438.

R. Kavanagh, T. Darcey, and D. Fender, "The Dimensionality of the Human Visual Evoked Scalp Potential," *Electroencephalogr. Clin. Neurophysiol*, vol. 40, pp. 633-644, 1976.

R.D. Kennett, "The Classification of Multi-Edge Shapes Using An Autoregressive Model and the Karhunen-Loeve Expansion," M.S.E.E Thesis, University of New Hampshire, 1985.

J. Kittler, "A Method of Feature Selection for High Dimensional Pattern Recognition Problems," *Proc. 2nd Int. Conf. on Pattern Recognition*, Denmark, 1974.

J. Kittler, "Feature Selection Methods Based on the Karhunen-Loeve Expansion," in *Pattern Recognition Theory and Application*, K.S. Fu and A.B. Whinston, Eds. Leyden, Holland:Noordhoft, 1977, pp. 61-74.

J. Kittler, "Feature Selection and Extraction," in *Handbook of Pattern Recognition and Image Processing*, T. Y. Young and K.S. Fu, Eds. Orlando, Florida: Academic, 1986, pp. 59-83.

J. Kittler and P.C. Young, "A New Approach to Feature Selection Based on the Karhunen-Loeve Expansion," *Pattern Recognition*, Vol. 5, pp. 335-352, 1973.

Z. Koles and R.B. Paranjape, "Topographic Mapping of the EEG: An Examination of Accuracy and Precision ", *Brain Topography*, (accepted for publication).

P. Nunez, *Electric Fields in the Brain: The Neurophysics of EEG*. New York: Oxford Univ. Press, 1981.

S. Periyalwar, S.T. Nugent, B.M. Horacek, "Application of Fourier and Eigenvector Analysis to the Classification of Integral Maps," in *Proceedings of CMBEC - 13 - CCGB*, Halifax, Nova Scotia, pp.183-184, 1987.

A. Rosenfeld, and A.C. Kak, *Digital Picture Processing*, New York: Academic,1982.

U. Schmiedl, D.A. Ortendahl, A.S. Mark, I. Berry, and L. Kaufman, "The Utility of Principal Component Analysis for the Image Display of Brain Lesions. A Preliminary, Comparative Study," *Mag. Res. in Med.*, Vol. 4, pp. 471-486, 1987.

M. Shimura and T. Imai, "Nonsupervised Classification Using the Principal Component," *Pattern Recognition*, Vol. 5, pp. 353-363, 1973.

W. Skrandies and D. Lehmann, "Spatial Principal Components of Multichannel Maps Evoked By Lateral Visual Half-Field Stimuli," *Electroencephalogr. Clin. Neurophysiol.*, vol. 54, pp. 662-667, 1982.

B.T. Smith, et al., *Matrix Eigensystem Routines - EISPACK GUIDE*, Springer Verlag, 1976.

J.T. Tou and R.P. Heydorn, "Some Approaches to Optimum Feature Extraction," *Computer and Information Sciences II*, New York: Academic, pp. 61-89, 1967.

H. Van hamme, M. Unser, E. Van Denhaute, K. Keymolen, J. Cornelis, "About the Digital Processing of Dynamic Infrared Thermography Image Sequences," in *Proceedings of CMBEC - 13 - CCGB*, Halifax, Nova Scotia, pp. 99-100,1987.